



## Is One's Own Ethnic Prejudice Always Subtle? The Inconsistency of Prejudice Endorsement and Prejudice Awareness Depends on Self-Related Egalitarian Standards and Motivations

Karolina Fetz & Tim Sven Müller

To cite this article: Karolina Fetz & Tim Sven Müller (2020) Is One's Own Ethnic Prejudice Always Subtle? The Inconsistency of Prejudice Endorsement and Prejudice Awareness Depends on Self-Related Egalitarian Standards and Motivations, Basic and Applied Social Psychology, 42:1, 1-28, DOI: [10.1080/01973533.2019.1689362](https://doi.org/10.1080/01973533.2019.1689362)

To link to this article: <https://doi.org/10.1080/01973533.2019.1689362>



© 2019 The Author(s). Published with  
license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 22 Nov 2019.



[Submit your article to this journal](#)



Article views: 325



[View related articles](#)



[View Crossmark data](#)

## Is One's Own Ethnic Prejudice Always Subtle? The Inconsistency of Prejudice Endorsement and Prejudice Awareness Depends on Self-Related Egalitarian Standards and Motivations

Karolina Fetz and Tim Sven Müller

Humboldt-Universität zu Berlin



### ABSTRACT


In light of current egalitarian norms, it seems intuitive to assume that people endorsing ethnically prejudicial beliefs are largely unaware of their xenophobic content. However, a cognitive consistency perspective would suggest that individuals with low anti-prejudice standards might care less if their opinions are prejudiced. Corroborating this idea, Study 1 ( $N = 919$ ) reveals that intra-individually the endorsement of prejudicial beliefs was negatively related to their evaluation as xenophobic (prejudice awareness), but more so among individuals with strong self-related egalitarian standards. Study 2 ( $N = 1,201$ ) replicates these findings and shows that egalitarian standards salience led to a stronger negative association between prejudice endorsement and awareness. This suggests that low-egalitarians are less suitable targets for awareness-raising components of anti-prejudice interventions.

In today's societies, prevailing egalitarian norms largely condemn ethnic prejudice<sup>1</sup> and anti-immigrant resentments, but nevertheless, social-psychological research and public opinion surveys disclose that people frequently hold prejudicial beliefs (e.g., Murray & Marx, 2013; Schneider, 2007; Strabac & Listhaug, 2008; Zick, Küpper, & Hövermann, 2011). When trying to reconcile these findings, it might seem intuitive to suppose that individuals who explicitly endorse ethnically prejudicial opinions, obviously do not recognize their beliefs as prejudicial in the first place, in assuming that it is nowadays rather aversive to perceive oneself as anti-egalitarian or prejudiced. In turn, one might infer that making individuals aware of this fact is a useful way to motivate them to reconsider their beliefs.<sup>2</sup> But can we really presume that people are generally unaware of the prejudicial nature of the opinions they personally endorse? Or might there also be individuals who do not care so much whether they hold ethnically prejudicial beliefs, because they do not have personal anti-prejudice standards?

Empirical studies investigating these questions are rather scarce. Social-psychological frameworks on contemporary forms of explicit ethnic prejudice (e.g.,

McConahay, 1986; Pettigrew & Meertens, 1995) suggest that along with a current trend towards egalitarianism it has become rather undesirable to appear as ethnically prejudiced. In turn, it has been argued that many people would therefore refrain from endorsing beliefs they clearly perceive as racist. This rationale would conversely imply that people are largely unaware of the prejudicial nature of those beliefs that they do endorse themselves, and that one's own ethnic prejudice is thus rather 'subtle to oneself'. This would, in other words, suggest that the endorsement of an ethnically prejudicial statement (prejudice endorsement) and the simultaneous evaluation of this belief as xenophobic or racist (prejudice awareness) are largely inconsistent with each other. This idea has been indirectly corroborated by research findings revealing that prejudice measures which receive higher endorsement levels are also rated as more socially acceptable or less reflective of racism (e.g., Manganelli Rattazzi & Volpato, 2003; McConahay, 1986), and that being confronted with one's own prejudiced responses has negative affective consequences (e.g., Fehr & Sassenberg, 2010; Monteith & Mark, 2005). However, previous research has not directly

**CONTACT** Karolina Fetz  [karolina.fetz@hu-berlin.de](mailto:karolina.fetz@hu-berlin.de)  Berlin Institute for Integration and Migration Research, Faculty of Humanities and Social Sciences, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.

 Supplemental material for this article can be accessed at [www.tandfonline.com/hbas](http://www.tandfonline.com/hbas).

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

investigated how people evaluate their own prejudicial opinions, in the first place.

Approaching this matter, the current article scrutinizes based on two studies—one correlational and one experimental—the inconsistency between prejudice endorsement and prejudice awareness, in taking an analytical and a methodological stance that explicitly focuses on intra-individual evaluation dynamics. Specifically, it investigates how prejudice endorsement and prejudice awareness are internally associated within individuals, i.e., in how far individuals evaluate different ethnically prejudicial beliefs they might endorse themselves as (not) xenophobic. Furthermore, drawing on a cognitive consistency perspective (e.g., Festinger, 1957; Gawronski, 2012), we examine whether prejudice endorsement and prejudice awareness are not generally inconsistent, but whether this might depend on individuals' standards and motivations condemning or condoning the personal endorsement of ethnic prejudice (Study 1 and 2) or on the salience of such egalitarian standards (Study 2). Especially highly egalitarian individuals (or individuals for whom egalitarianism is made salient) might more strongly critically monitor their own beliefs and might only endorse prejudicial beliefs to the extent that they are unaware of their xenophobic nature, as it would otherwise collide with personal (or salient) anti-prejudice standards.

These studies seek to provide insights into cognitive processes behind the subtlety of ethnic prejudice and to further illuminate which individuals are the most suitable target group for components of interventions aimed at combating ethnic prejudice by raising awareness of own prejudicial beliefs. Moreover, in drawing on data from two large German population samples, we aim to provide a more representative picture of these prejudice-related cognitive mechanisms, which we consider especially important with regard to current debates on the generalizability and contextualization of social-psychological research findings (Pettigrew, 2018).

### The subtlety of contemporary ethnic prejudice

The idea that many people would rather not endorse beliefs they define as racist or reflective of ethnic prejudice is implied within a variety of social-psychological approaches on contemporary forms of explicit ethnic prejudice, which start from the common assumption that ethnic prejudice now appears in subtler guises that circumvent egalitarian anti-prejudice norms in today's societies (e.g., Henry &

Sears, 2002; Kinder & Sears, 1981; McConahay, 1986; McConahay, Hardee, & Batts, 1981; Pettigrew & Meertens, 1995). One of the main goals of these frameworks, such as modern racism (McConahay, 1986; McConahay et al., 1981), and subtle prejudice (Pettigrew & Meertens, 1995), was the development of new self-report measures that are less reactive, in containing items which are less readily perceived to be in conflict with norms condemning ethnic prejudice. Higher levels of endorsement for these new prejudice measures—compared to a lower endorsement of more old-fashioned ethnically prejudicial beliefs—were thereby taken as an indicator of their subtlety. Behind this inference lies the assumption—which might also be conceptualized as linked to the basic tenets of cognitive dissonance theory (Festinger, 1957)—that many people would nowadays want to maintain an egalitarian, unprejudiced self-image for themselves (but also in front of others), and would thus rather not endorse 'blatant' beliefs on self-report (i.e., explicit) measures, they evaluate as racist (e.g., McConahay, 1986, p. 100; Pettigrew & Meertens, 1995, p. 73; see also Pearson, Dovidio, & Gaertner, 2009, regarding the discussion on the desirability of the maintenance of an egalitarian self-image related to more implicit forms of racism). This conversely implies that those ethnically prejudicial beliefs which many individuals do explicitly endorse are rather 'subtle', in being largely not perceived by them as racist or xenophobic. In that vein, this rationale of approaches on subtle forms of explicit ethnic prejudice to some degree suggests an inconsistency of prejudice endorsement and prejudice awareness, which implicitly relies on the supposition of an internal self-monitoring or balancing process, whereby many individuals indicate their (dis)agreement with a prejudicial statement also based on their concurrent evaluation of this statement as (not) being reflective of ethnic prejudice.

Yet, only a limited number of studies have directly investigated this rationale. Crandall, Eshleman, and O'Brien (2002), for instance, found a nearly perfect, positive relationship between endorsement and social acceptability ratings for a variety of different social prejudices, indicating that the higher a prejudice is on average endorsed, the more it is regarded as socially acceptable. Similarly, other studies showed that items belonging to measures of contemporary ethnic prejudice, which commonly receive higher levels of endorsement compared to old-fashioned prejudice scales, were indeed subtler, in being rated as more socially acceptable or less racist than old-fashioned prejudice items (Manganelli Rattazzi & Volpato, 2003;

McConahay, 1986; McConahay et al., 1981; Pettigrew & Meertens, 1996). Taken together, these studies indeed indicate an inconsistency between prejudice endorsement and prejudice awareness, mirrored by their negative association, *on the level of specific prejudicial statements*. Thus, they show from a social-normative perspective that prejudicial statements that are on average more strongly endorsed are on average also subtler in being perceived as more socially acceptable or less racist.

However, in operating at the item level, these studies remain inconclusive on whether prejudice endorsement and prejudice awareness are inconsistent on an individual level, i.e., whether a higher endorsement of a prejudicial statement is similarly associated with a lower degree of awareness of its xenophobic content *within respondents*. Hence, they do not disclose whether one's own prejudice is also rather 'subtle to oneself', in the sense that individuals, for themselves, do mostly not evaluate the beliefs they personally endorse as xenophobic or racist. Such an analytical shift to an intra-individual perspective is however crucial, especially with regard to interventions that are based on the idea that making people aware of their own (ethnically) prejudicial beliefs is a useful means to combat such opinions. The only study that has to our knowledge touched upon this question focused on sexist prejudice. Swim, Mallett, Russo-Devosa, and Stangor (2005) found that individuals with higher endorsement levels on different sexism measures, showed, on average, lower awareness for the prejudicial nature of the same measures, in judging them less sexist. However, in employing aggregated scale scores per respondent within this study, it still remains uncertain to which degree the individual endorsement of a specific prejudicial statement (or prejudice item) is specifically related to the degree to which that same statement is evaluated as prejudicial.

In sum, these findings provide first indirect support for an inconsistency between prejudice endorsement and prejudice awareness, empirically reflected by their negative association. This might tentatively suggest that individuals indeed engage in a critical self-monitoring of own beliefs, explicitly endorsing ethnically prejudicial only to the degree that they are unaware of their xenophobic nature. Nevertheless, especially with regard to ethnic prejudice, there is a lack of research investigating whether the endorsement of a prejudicial statement is really unlikely to coincide with the evaluation of the same statement as xenophobic, by taking an analytical as well as a methodological stance that directly addresses this idea on

an intra-individual level. This is however crucial in order to open the black box of cognitive processes behind the subtlety of one's own ethnic prejudice.

### **The affective costs of prejudice awareness depend on personal egalitarian standards**

Further indirect evidence for the idea that one's own ethnic prejudice might be generally rather subtle to oneself and that people are likely to be rather unaware of the prejudicial nature of those beliefs they personally endorse stems from studies that have examined the fruitfulness of combating prejudice by inducing awareness of own prejudice-related responses. This line of research can be conceptualized as exploiting the fact that people strive to achieve consistency between different values or beliefs they hold as well as behaviors they exhibit, as it has been put forward within cognitive dissonance theory (Festinger, 1957)—and in a comparable manner also in related theoretical frameworks such as self-discrepancy theory (Higgins, 1987) or symbolic self-completion theory (Wicklund & Gollwitzer, 1981, 1982), which propose that individuals seek to reduce discrepancies between their current self-perception and important, aspired self-definitions or states of the self they seek to attain. Within this research, it is suggested that confronting people with the fact that own responses are reflective of (ethnic) prejudice and potentially violate personal egalitarian standards can cause negative self-directed emotions. In turn, this negative affective state is considered to act as a motivator for individuals to control prejudiced responses. Within some of these studies, participants are, for instance, provided with feedback of having supposedly exhibited prejudiced responses or discriminatory behavior (e.g., Fehr & Sassenberg, 2010; Monteith, 1993), while in other studies, people are asked to critically self-reflect how they should and would respond in certain prejudice-related scenarios (e.g., Devine et al., 1991; Voils, Ashburn-Nardo, & Monteith, 2002; Zuwerink, Devine, Monteith, & Cook, 1996). Overall, it is a stable empirical finding that pinpointing (supposed) prejudiced responses leads to negative self-directed affect, i.e., in Allport's (1954) or Festinger's (1957) words 'compunction' or 'cognitive dissonance', only (or more strongly) for low-prejudiced individuals and people with strongly internalized anti-prejudice standards (e.g., Fehr & Sassenberg, 2010; see also Monteith & Mark, 2005, for a review), causing especially these individuals to initiate prejudice-related self-regulation mechanisms (e.g., Fehr & Sassenberg, 2010; Monteith, 1993). This is in line with

cognitive dissonance theory (Festinger, 1957), since prejudiced reactions should represent a discrepancy or inconsistency with personal values only for individuals with strong egalitarian anti-prejudice standards.

Taken together, although these findings provide an important starting point for anti-prejudice interventions, in revealing that making people aware of their own prejudicial responses can be a useful means to combat (ethnic) prejudice, also this line of previous research did not directly examine how people evaluate their own prejudiced responses—especially their consciously endorsed, explicit prejudicial opinions—for themselves. More specifically, and with regard to the current research question, these studies do not disclose, whether prejudice endorsement and prejudice awareness are, in the first place, inconsistent, in the sense that most people are initially and prior to potential interventions, largely unaware of the (ethnically) prejudicial nature of their own beliefs. Of course, the empirical results revealing the aversive emotional consequences of confronting people with their (alleged) prejudice-related responses can, on the one hand, be interpreted as indirect evidence that many individuals are indeed unaware that the prejudicial beliefs they do personally endorse are actually reflective of (ethnic) prejudice. Yet, on the other hand, the finding that such interventions cause negative affect particularly (or only) for highly egalitarian individuals also suggests that simultaneous prejudice endorsement and prejudice awareness might not be equally inconsistent for all people. Especially individuals with standards and values that condemn the personal endorsement of (ethnic) prejudice might more strongly monitor their explicit beliefs as potentially prejudiced and might initially be more strongly convinced that their own ethnically prejudicial beliefs are not xenophobic, in assuming that they would not endure the otherwise resulting negative affective state of cognitive dissonance.

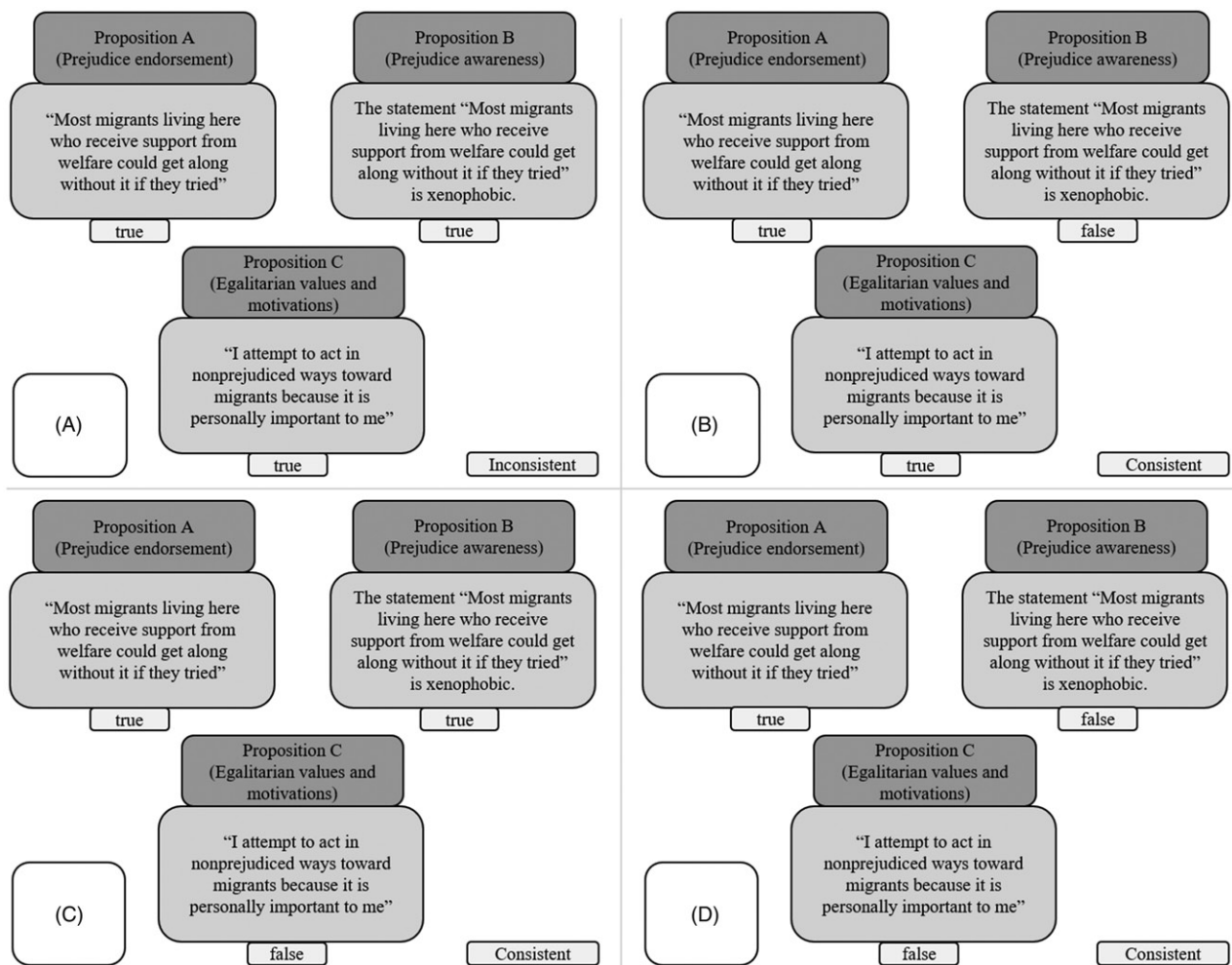
### **A cognitive consistency approach to the subtlety of one's own ethnically prejudicial beliefs**

Expanding this idea, the studies presented here focus on intra-individual cognitive processes and seek to directly investigate in how far people privately do or do not evaluate the ethnically prejudicial beliefs they personally and explicitly endorse as xenophobic. Moreover, in drawing on a cognitive consistency perspective the current research examines whether the supposed inconsistency between prejudice

endorsement and prejudice awareness, which would be reflected by a negative intra-individual association, might be determined by personal values, standards or motivations that condemn or condone the own endorsement of ethnic prejudice (Study 1 and 2) or the salience of such standards (Study 2).

Following the rationale put forward by Gawronski and colleagues (e.g., Gawronski, 2012; Gawronski & Bodenhausen, 2014; Gawronski & Brannon, 2019; Gawronski, Peters, Brochu, & Strack, 2008; Gawronski & Strack, 2004), which was based on cognitive dissonance theory (Festinger, 1957), we apply a cognitive consistency framework to prejudice-related belief systems and regard cognitive (in)consistency as a propositional process: Two or more propositional beliefs that are logically related to each other and are (subjectively perceived as) true or false can be (in)consistent with each other, potentially causing the negative affective state of cognitive dissonance. Transferring this idea to the context of the current study, one could regard the explicit endorsement of an ethnically prejudicial statement (prejudice endorsement) and the explicit evaluation of the same statement as xenophobic (prejudice awareness) as two inter-related propositional beliefs that can be subjectively evaluated as true or false. For instance, individuals could or could not endorse an ethnically prejudicial belief (e.g., “Most migrants living here who receive support from welfare could get along without it if they tried”, i.e., an item adapted from Pettigrew and Meertens’ (1995) *Blatant Prejudice Scale*), i.e., assign it a positive or negative truth value. Concurrently, individuals can evaluate that same belief as xenophobic or not xenophobic, and could thus, in this example, judge the proposition “The statement »Most migrants living here who receive support from welfare could get along without it if they tried« is xenophobic” as true or false. Although one might intuitively regard the simultaneous agreement with these two statements, indicating concurrent prejudice endorsement and prejudice awareness, as contradictory, these two beliefs are from the perspective of propositional logic by themselves neither consistent nor inconsistent with each other, regardless of the truth values assigned to them. However, one could argue, that their logical relationship (or rather ‘psycho-logical’ relationship, see e.g., Gawronski & Brannon, 2019) is defined by personal values, standards or motivations which specify in how far being personally unprejudiced is actually important to an individual. One could hereby, on the one hand, think of one’s internal motivation to respond without prejudice (IM; Plant & Devine, 1998), a standard to avoid





**Figure 1.** Schematic representation of prejudice-related belief systems and their (in)consistency comprising the endorsement of an ethnically prejudicial belief, the awareness that this belief is reflective of ethnic prejudice, and the endorsement of egalitarian values and motivations. *Note.* Panel A represents an inconsistent belief system; Panels B, C, and D represent consistent belief systems. This representation was developed based on the work of Gawronski et al. (2008).

prejudice derived from internalized anti-prejudice norms, the general egalitarian self-perception (ES), and humanitarianism-egalitarianism (HE; Katz & Hass, 1988), a value orientation emphasizing equality and social justice, as individual difference variables condemning the private endorsement of ethnic prejudice. On the other hand, social dominance orientation (SDO; e.g., Ho et al., 2012; Ho et al., 2015; Pratto, Sidanius, Stallworth, & Malle, 1994; Sidanius, Pratto, & Bobo, 1996), as an ideological preference for group-based hierarchies, external motivation to respond without prejudice (EM; Plant & Devine, 1998), a standard to respond unprejudiced due to external social pressure,<sup>3</sup> and protestant work ethic (PWE; Katz & Hass, 1988), a value orientation emphasizing individual responsibility for one's success or failure, would be individual difference variables which condone the private endorsement of ethnic prejudice. Treating such standards and motivations as

propositional beliefs themselves (e.g., "I attempt to act in nonprejudiced ways toward migrants because it is personally important to me", i.e., an item adapted from Plant and Devine's (1998) IM scale), their endorsement or rejection, respectively, in combination with the truth values assigned to the other two propositions could result in an inconsistent belief system.<sup>4</sup> Approaching the question of whether individuals are likely to be unaware of their own ethnic prejudice from a cognitive consistency perspective, Figure 1 schematically depicts four possible intra-individual belief systems, starting from the assumption that an individual endorses an ethnically prejudicial belief (Proposition A is true, prejudice endorsement). If an individual at the same time endorses standards that condemn personal prejudice endorsement (Proposition C is true), evaluating the own prejudicial belief as xenophobic (Proposition B is true, prejudice awareness) should collide with personal values,

resulting in cognitive inconsistency and ultimately cognitive dissonance (Panel A). Thus, if highly egalitarian individuals endorse an ethnically prejudicial statement, the internal belief system is only consistent if they evaluate this ethnically prejudicial belief at the same time as not xenophobic (Proposition B is false, no prejudice awareness; Panel B). On the other hand, for individuals who reject standards that condemn personal prejudice endorsement (Proposition C is false), whether or not one privately evaluates the prejudicial beliefs one endorses as xenophobic should not collide with own values and should thus not affect the consistency of the belief system (Panel C and D). The same logic would apply for individuals rejecting or endorsing values and standards, respectively, that condone personal prejudice endorsement.

Thus, from this cognitive consistency perspective, one could argue that the (logical) inconsistency of prejudice endorsement and prejudice awareness depends on individual (anti-)egalitarian values and motivations that condemn or condone personal endorsement of ethnic prejudice, and that their intra-individual association might thus be moderated by such standards. On the one hand, individuals with strong standards condemning (and low standards condoning) personal prejudice endorsement might more strongly engage in a self-monitoring of their own beliefs as potentially prejudiced, as this is relevant for the consistency of their belief system. Hence, when indicating their endorsement with a statement, they might more strongly simultaneously consider whether this utterance is reflective of ethnic prejudice. In turn, these individuals might only explicitly endorse prejudicial beliefs to the extent that they are unaware of their xenophobic nature and might more strongly evaluate those prejudicial beliefs they do endorse as not xenophobic (Panel B), assuming that they would not endure the cognitive dissonance that would result from the otherwise inconsistent belief system. Thus, since for highly egalitarian individuals prejudice endorsement and prejudice awareness are indeed (logically) inconsistent with each other, one could expect them to exhibit a stronger negative intra-individual relationship between prejudice endorsement and prejudice awareness. In contrast, one could argue that people with low standards condemning (and high standards condoning) personal prejudice endorsement might less strongly self-monitor their beliefs and consider whether they think that the beliefs they privately endorse are xenophobic, as this is less relevant for the consistency of their belief system. Thus, since for these less egalitarian individuals the explicit

endorsement of a prejudicial statement and its simultaneous evaluation as xenophobic are not logically inconsistent with each other and should not cause cognitive dissonance, one could expect that for them, prejudice endorsement and prejudice awareness would be intra-individually more unrelated.

Overall, such potential differences in the intra-individual association of prejudice endorsement and prejudice awareness could serve as an important indicator for the suitability of those components of anti-prejudice interventions that aim to combat ethnic prejudice by raising awareness of own prejudicial beliefs, as these can only be a reasonable means for the reduction of explicit ethnic prejudice in the real world for individuals who are actually unaware of the prejudicial nature of their beliefs beforehand.

## Overview

In sum, it is the goal of the current research to examine within two studies, whether individuals' own ethnic prejudice is generally subtle to oneself, in the sense that people are mostly unaware of the prejudicial nature of the beliefs they personally and explicitly endorse. More specifically, we explore whether prejudice endorsement and prejudice awareness are inconsistent in investigating their *intra-individual* association, i.e., in how far the endorsement of different ethnically prejudicial statements is related to the evaluation of the same statements as xenophobic within individuals. Employing a cognitive consistency perspective, we further seek to examine whether this relationship between prejudice endorsement and prejudice awareness might depend on individual (anti-)egalitarian standards (Study 1 and 2) and might thus also be affected by the temporary salience of such standards (Study 2).

## Study 1

In Study 1, we investigate based on correlational survey data, how prejudice endorsement and prejudice awareness are intra-individually associated and in how far this relationship is determined by individual (anti-)egalitarian standards and motivations condemning or condoning the personal endorsement of ethnic prejudice. Based on previous findings, we generally expect a negative association of prejudice endorsement and prejudice awareness and predict that the more individuals endorse an ethnically prejudicial belief, the less they would evaluate this belief as xenophobic. However, from a cognitive consistency perspective, we hypothesize that this intra-individual association is moderated by individual

(anti-)egalitarian standards. Specifically, we predict that with increasing values, standards and motivations that condemn, i.e., IM, ES, and HE, and decreasing standards that condone the private endorsement of ethnic prejudice, i.e., SDO, EM, and PWE, individuals would more strongly engage in a self-monitoring of their own beliefs as potentially prejudiced and exhibit a stronger negative association between prejudice endorsement and prejudice awareness. Conversely, we expect that with lower standards that condemn and higher standards that condone the private endorsement of ethnic prejudice, prejudice endorsement and prejudice awareness would become less negatively associated and more unrelated within respondents.

## Method

### Participants and procedure

Data for the current study were collected as part of a larger telephone survey on attitudes towards migration and ethnic minorities among adults in Germany (Fetz & Kroh, 2019), conducted by a professional survey institute. Telephone surveys have the advantage that they allow to draw a probability sample of the general population, which increases the external validity of the study compared to other survey modes with convenience samples.<sup>5</sup> Also, compared to face-to-face interviews that permit probability sampling as well, telephone surveys are associated with lower social desirability bias (e.g., Schwarz, Strack, Hippler, & Bishop, 1991).

With regard to the measures relevant for the current study, the survey procedure was as follows: In the first part of the interview, respondents indicated their agreement with different prejudice items that varied in the strength of their prejudicial content (prejudice endorsement). Then, socio-demographic characteristics were obtained. After that, participants were informed that they would now be exposed to the same prejudice items they had evaluated beforehand, but should now give a different judgment, indicating in how far they evaluate these statements as xenophobic (prejudice awareness). Individual difference variables were assessed in the last part of the interview. Participation in the survey was entirely voluntary, respondents received no compensation for participation and could quit the interview at any time.

After excluding 60 cases with missing values on the individual difference variables in order to keep the sample constant across all statistical models, and one participant with only missing values on the main outcome variable, the final random population sample

for our analyses comprised 919 participants (465 women, 452 men, 2 diverse/no indication;  $M_{\text{age}} = 49.54$  years,  $SD = 16.47$ , age range: 18–87 years).

## Measures

### Level 1 variables (item ratings nested within respondents)

**Prejudice endorsement.** Two items each were selected from the Subtle and Blatant Prejudice Scale (Pettigrew & Meertens, 1995), translated into German (following the translation provided by Zick, 1997) and rephrased to assess prejudice towards migrants in Germany. Items from both scales, thus differing in the strength of the prejudicial content, were chosen in order to achieve sufficient within-subject variation in item ratings, which was essential for the purpose of the study. Additionally, the item selection was based on the premise that the wording needed to be logically sound for endorsement as well as awareness ratings (see below). The four employed items were: “Migrants have jobs that the Germans should have”, “Most migrants living here who receive support from welfare could get along without it if they tried”, “It is just a matter of some people not trying hard enough. If migrants would only try harder they could be as well off as German people”, and “Migrants living here teach their children values and skills different from those required to be successful in Germany”. Participants indicated their agreement with the prejudice items on a 1 (*completely disagree*) to 5 (*completely agree*) scale. Mean endorsement ratings and standard deviations for the four items were  $M = 1.75$ ,  $SD = 1.05$ ;  $M = 2.01$ ,  $SD = 1.14$ ;  $M = 2.56$ ,  $SD = 1.30$ ;  $M = 2.74$ ,  $SD = 1.21$ , respectively.

**Prejudice awareness.** Afterwards, participants indicated for the same four prejudice items, in how far they evaluate these as xenophobic on a 1 (*not xenophobic at all*) to 5 (*very xenophobic*) scale.<sup>6</sup> Mean awareness ratings and standard deviations for the four items were  $M = 3.91$ ,  $SD = 1.16$ ;  $M = 3.70$ ,  $SD = 1.20$ ;  $M = 3.51$ ,  $SD = 1.31$ ;  $M = 3.07$ ,  $SD = 1.19$ , respectively.

Endorsement and awareness ratings for the prejudice items were treated as separate observations clustered within individuals in our analyses and were thus not averaged per respondent.

### Level 2 (individual difference) variables

For all individual difference measures, respondents indicated their agreement with the respective items on a 1 (*completely disagree*) to 5 (*completely agree*) scale.



**Internal and external motivation to respond without prejudice.** Participants' motivations to respond without prejudice were assessed with two items each from the IM (Plant & Devine, 1998; e.g., "I attempt to act in nonprejudiced ways toward migrants because it is personally important to me") and the EM scale (e.g., "I attempt to appear nonprejudiced toward migrants in order to avoid disapproval from others"). Items were translated into German and rephrased to assess motivations to control prejudice towards migrants. The two IM ( $M=4.53$ ,  $SD=0.74$ ,  $\alpha=.75$ ) and EM ( $M=3.24$ ,  $SD=1.25$ ,  $\alpha=.55$ ) items were averaged for each individual.<sup>7</sup>

**Egalitarian self-perception.** Respondents' ES was measured with one item ("I am a person to whom equality of all groups in our society is important";  $M=4.51$ ,  $SD=0.79$ ).

**Social dominance orientation.** SDO was assessed with a German translation of the 8-item SDO<sub>7(s)</sub>-Scale (Ho et al., 2015; e.g., "An ideal society requires some groups to be on top and others to be on the bottom";  $M=1.89$ ,  $SD=0.64$ ,  $\alpha=.69$ ).

**Humanitarianism-egalitarianism.** HE was assessed with six items from Doll and Dick's (2000) German adaptation of the HE scale (Katz & Hass, 1988; e.g., "Those who are unable to provide for their basic needs should be helped by others";  $M=4.28$ ,  $SD=0.58$ ,  $\alpha=.71$ ).

**Protestant work ethic.** Endorsement of PWE was measured with a German adaptation of the four items employed by Levin, Sidanius, Rabinowitz, and Federico (1998; e.g., "Most people who don't get ahead should not blame the system; they really have only themselves to blame";  $M=2.59$ ,  $SD=0.76$ ,  $\alpha=.70$ ).

### Data structure and analytic strategy

The current study aims to assess the intra-individual association of the endorsement of the four prejudice items of different strength and the evaluation of the same items as xenophobic (prejudice awareness), and in how far this relationship depends on personal standards and motivations. In order to be able to analyze the covariation of prejudice endorsement and prejudice awareness within individuals, we transformed the data into longitudinal format, with each row representing the endorsement and awareness

ratings for the different prejudice items as paired repeated measurements clustered within respondents. This resulted in a two-level hierarchical data structure: After the exclusion of missing observations using listwise deletion, 3,588 paired endorsement and awareness ratings (level 1) were nested within 919 respondents (level 2). Employing multilevel analyses (Snijders & Bosker, 2012), we first calculated multilevel models (maximum likelihood estimation) with respondent-specific intercepts and slopes in order to assess in how far the endorsement of a prejudice item predicted the awareness rating for the same item within participants. Prejudice awareness was entered as the outcome and prejudice endorsement as the predictor variable in the models.<sup>8</sup> In order to examine how the intra-individual association between prejudice endorsement and prejudice awareness (i.e., the individual slope) varied between individuals depending on respondents' values and motivations, we then estimated multilevel models including cross-level interactions with respondent characteristics (level 2 variables; i.e., IM, ES, SDO, EM, HE, PWE). The covariance between the random effects was not restricted. Beforehand, the outcome and all predictor variables were z-standardized on their respective level of analysis. All multilevel models were estimated using the xtmixed command for multilevel regression models in Stata (Version 15.1; StataCorp, 2017).

### Results

Exploring the hierarchical data structure, we first calculated an unconditional model with prejudice awareness as the outcome variable and no explanatory variables being entered (Model 0; see Table 1). Results from this null model confirmed that an adequate amount of variance existed on both levels. The estimated random effects parameters showed that respondents sufficiently varied in their awareness ratings provided for the different prejudice items (variance within respondents), and participants substantially differed from each other in their mean level of prejudice awareness across these items (variance between respondents). The intraclass correlation (ICC) coefficient indicated that a high proportion of 43.76% of the variance in awareness ratings could be attributed to the respondent level, i.e., participants tended to evaluate the four prejudice items quite similarly.

In the next step, we calculated multilevel models with prejudice awareness as the outcome variable and the endorsement for the prejudice items as the main

**Table 1.** Multilevel regression analyses of the intra-individual effect of prejudice endorsement (level 1) and its cross-level interaction with individual difference characteristics (level 2) on prejudice awareness.

Parameters and variables	Model 0 <i>b</i> ( <i>SE</i> )	Model 1 <i>b</i> ( <i>SE</i> )	Model 2 <i>b</i> ( <i>SE</i> )	Model 3 <i>b</i> ( <i>SE</i> )	Model 4 <i>b</i> ( <i>SE</i> )	Model 5 <i>b</i> ( <i>SE</i> )	Model 6 <i>b</i> ( <i>SE</i> )	Model 7 <i>b</i> ( <i>SE</i> )	Model 8 <i>b</i> ( <i>SE</i> )
<b>Fixed</b>									
Level 1									
Intercept	.004 (.025)	.138 (.029)	.135 (.029)	.139 (.029)	.138 (.029)	.132 (.029)	.144 (.029)	.141 (.029)	.134 (.029)
Prejudice endorsement		-.347 (.017)	-.329 (.018)	-.335 (.017)	-.331 (.018)	-.350 (.017)	-.330 (.018)	-.342 (.017)	-.322 (.018)
Level 2									
IM			.204 (.022)	.207 (.021)					.113 (.025)
ES					-.153 (.022)				.129 (.024)
SDO									-.049 (.024)
EM						-.007 (.022)			-.021 (.021)
HE							.158 (.022)		.038 (.025)
PWE								-.061 (.022)	
<b>Cross-level interactions:</b>									
Prejudice endorsement x									
IM			-.051 (.015)	-.051 (.015)					-.038 (.017)
ES									-.038 (.017)
SDO									.009 (.018)
EM					.033 (.016)	.048 (.016)			.045 (.017)
HE							-.020 (.016)		.010 (.018)
PWE								-.009 (.016)	
<b>Random</b>									
Random intercept ( $\sigma_0^2$ )	Estimate .437 (.027)	Estimate .302 (.021)	Estimate .270 (.019)	Estimate .265 (.019)	Estimate .283 (.020)	Estimate .300 (.021)	Estimate .283 (.020)	Estimate .300 (.021)	Estimate .248 (.018)
Random slope ( $\sigma_1^2$ )		.058 (.010)	.055 (.010)	.057 (.010)	.058 (.010)	.057 (.010)	.056 (.010)	.057 (.010)	.053 (.010)
Covariance ( $\sigma_{01}^2$ )		-.023 (.010)	-.015 (.009)	-.015 (.009)	-.018 (.009)	-.023 (.010)	-.019 (.009)	-.023 (.010)	-.009 (.009)
Residual variance	.561 (.015)	.413 (.013)	.413 (.013)	.412 (.013)	.412 (.013)	.413 (.013)	.412 (.013)	.413 (.013)	.412 (.012)
ICC	.438								
BIC	9,414.747	8,583.551	8,514.710	8,502.809	8,548.854	8,591.311	8,547.358	8,591.571	8,514.511
Deviance	9,390.191	8,509.883	8,424.671	8,412.770	8,458.816	8,501.272	8,457.319	8,501.532	8,358.989

Based on  $N = 3,588$  item ratings clustered within  $N = 919$  participants. Predictor and outcome variables were z-standardized on their respective level of analysis. Models 1–8 included a categorical control variable for the prejudice items. The BIC and deviance ( $-2$  log-likelihood) statistics give an indication of model fit, with lower values representing a better fit.

predictor variable on level 1. The ‘raw’, average intra-individual effect of prejudice endorsement on prejudice awareness, based on a model without any control variables, was negative—as expected—and of moderate size ( $b = -.412$ ,  $SE = .017$ ; see Table 2 for an overview on the simple intra-individual effects of prejudice endorsement on prejudice awareness estimated for Study 1 and Study 2). We then included a categorical control variable for the specific prejudice items in the models for the main analyses, which are reported hereafter (Models 1–8; see Table 1). Further corroborating our hypothesis, results from Model 1 revealed a negative main effect of prejudice endorsement on prejudice awareness ( $b = -.347$ ,  $SE = .017$ ), i.e., a moderate negative intra-individual association between prejudice endorsement and awareness. Thus, the more individuals endorsed a prejudicial belief, the less they evaluated the same belief as xenophobic. The random slope estimate indicated that this negative association between prejudice endorsement and prejudice awareness varied between respondents with a standard deviation of .241, i.e., within a  $\pm 1.96$  SD range of  $[-.819; .125]$ . Thus, the intra-individual association of endorsement and awareness ratings was estimated to range from a stronger negative relationship to a complete unrelatedness between respondents.

In a second step, we examined in how far the intra-individual relationship between prejudice endorsement and prejudice awareness might differ between respondents, in being moderated by individual difference (level 2) variables. For this purpose, we first calculated separate multilevel models with prejudice awareness as the outcome and prejudice endorsement as the predictor variable on level 1, now including the respective individual difference (level 2) variables and most importantly the cross-level interaction terms between these individual difference variables and prejudice endorsement (Models 2–7). In line with our hypotheses, results from Models 2–6 revealed small interaction effects between endorsement ratings and IM, ES, SDO, EM, and HE respectively, on awareness ratings. Within respondents with higher IM, an increase in the endorsement of a prejudice item was more strongly related to a decrease in prejudice awareness for the same item ( $b = -.051$ ,  $SE = .015$ ). Similarly, with higher individual levels of ES, the intra-individual association between prejudice endorsement and prejudice awareness became more pronounced and negative ( $b = -.051$ ,  $SE = .015$ ). Consequently, at lower levels of IM and ES, the endorsement of prejudice items less strongly

**Table 2.** Overview of simple intra-individual effects of prejudice endorsement on prejudice awareness (Study 1 and Study 2)

	Study 1			Study 2 (Complete sample)				Study 2 (Adjusted sample)			
	Overall sample $Nn = 919$ $b$ (SE)	Control condition $n = 492$ $b$ (SE)	Egalitarian standard failure $n = 450$ $b$ (SE)	Egalitarian standard affirmation $n = 259$ $b$ (SE)	Overall sample $n = 888$ $b$ (SE)	Control condition $n = 463$ $b$ (SE)	Egalitarian standard failure $n = 257$ $b$ (SE)	Egalitarian standard affirmation $n = 168$ $b$ (SE)	Overall sample $n = 888$ $b$ (SE)	Control condition $n = 463$ $b$ (SE)	Egalitarian standard failure $n = 257$ $b$ (SE)
Prejudice endorsement	-.412 (.017)	-.493 (.011)	-.475 (.018)	-.511 (.017)	-.494 (.024)	-.535 (.012)	-.582 (.020)	-.595 (.025)	-.535 (.012)	-.482 (.017)	-.582 (.020)

Coefficients are estimated based on separate multilevel regression models for the respective (sub)samples. Models contained no control variables. Prejudice endorsement was z-standardized on its respective level of analysis and on the respective sample referred to (i.e., Study 1 (complete sample), Study 2 (complete or adjusted sample)).

negatively predicted the awareness ratings for the same items within individuals. In contrast, with increasing SDO, endorsement and awareness ratings were intra-individually less negatively associated and more unrelated ( $b = .033$ ,  $SE = .016$ ). Similarly, with higher EM, an increase in the endorsement of a prejudice item was less strongly associated and more unrelated to the awareness ratings for the same item ( $b = .048$ ,  $SE = .016$ ). In turn, at lower levels of SDO and EM, the endorsement of prejudice items was more strongly negatively associated with the awareness ratings for the same items within respondents. A very small interactive effect was found for HE ( $b = -.020$ ,  $SE = .016$ ). In order to make the size of these cross-level interaction effects more tangible, simple slope analyses were conducted that illustrate the difference in the effect of prejudice endorsement on prejudice awareness for individuals at lower ( $-1$  SD) as well as higher ( $+1$  SD) levels of IM, ES, SDO, EM, and HE. The association between prejudice endorsement and prejudice awareness differed for individuals who are low ( $b = -.279$ ,  $SE = .022$ ) and high ( $b = -.380$ ,  $SE = .024$ ) in IM by .101, for individuals who are low ( $b = -.284$ ,  $SE = .022$ ) and high ( $b = -.386$ ,  $SE = .024$ ) in ES by .102, for individuals who are low ( $b = -.365$ ,  $SE = .025$ ) and high ( $b = -.298$ ,  $SE = .023$ ) in SDO by .067, for individuals who are low ( $b = -.398$ ,  $SE = .025$ ) and high ( $b = -.302$ ,  $SE = .023$ ) in EM by .096, and for individuals who are low ( $b = -.309$ ,  $SE = .023$ ) and high ( $b = -.350$ ,  $SE = .024$ ) in HE by .041. These simple slope analyses reveal that the association between prejudice endorsement and prejudice awareness remains negative at both, low and high levels of IM, ES, SDO, EM and HE, but that its strength varies at different levels of the moderators. Contrary to our hypotheses, the interactive effect of prejudice endorsement and PWE on prejudice awareness was negligible ( $b = -.009$ ,  $SE = .016$ ; Model 7). Taken together, the results of Models 2–6 suggest that with increasing levels of IM, ES, and HE, and with decreasing levels of SDO and EM, individuals more strongly evaluated the prejudicial beliefs they endorsed themselves as not xenophobic. Conversely, for individuals with lower IM, ES, and HE, and higher SDO and EM, the endorsement of a prejudicial statement was less strongly negatively associated and more unrelated to its evaluation as xenophobic.

In the final step, we calculated a full model (Model 8), with prejudice awareness as the outcome variable, and including as predictor variables the endorsement ratings (level 1) as well as all individual difference (level 2) variables and their interaction terms with

prejudice endorsement that proved to be meaningful moderators in the separate analyses.<sup>9</sup> Interaction effects of endorsement ratings and IM, EM and ES, respectively, remained largely uniform. With increasing IM and ES, the intra-individual association between prejudice endorsement and prejudice awareness became more pronounced and negative ( $b = -.038$ ,  $SE = .017$ ; and  $b = -.038$ ,  $SE = .017$ , respectively). Also, with higher levels of EM, endorsement and awareness ratings were less strongly negatively associated, and more unrelated within individuals ( $b = .045$ ,  $SE = .017$ ). However, the interactive effect of prejudice endorsement and SDO and HE vanished in the full model ( $b = .009$ ,  $SE = .018$ ; and  $b = .010$ ,  $SE = .018$ , respectively). Compared to the first model including only prejudice endorsement as a (level 1) predictor variable for prejudice awareness ( $BIC = 8,583.551$ ), the full model exhibited an increased fit with the data ( $BIC = 8,514.511$ ). Additionally, the inclusion of the individual difference (level 2) variables and their cross-level interaction effects with prejudice endorsement (level 1) contributed to the explanation of random slope variance, that was reduced by 8.62% from .058 (Model 1) to .053 in the full model (Model 8).<sup>10</sup>

## Discussion

In line with our hypotheses, results from Study 1 overall showed a moderate, negative intra-individual association between prejudice endorsement and prejudice awareness. The more individuals agreed with an ethnically prejudicial statement, the less they tended to evaluate this statement as xenophobic. In line with our predictions derived from a cognitive consistency perspective, the inconsistency of prejudice endorsement and prejudice awareness seems to depend on personal standards, condemning or condoning the private endorsement of ethnic prejudice. Indeed, our findings revealed that the negative association of prejudice endorsement and prejudice awareness was more pronounced for individuals with higher egalitarian and lower anti-egalitarian standards and motivations. With higher levels of internal motivation to respond without prejudice (IM), egalitarian self-perception (ES) and humanitarianism-egalitarianism (HE), and with lower levels of social dominance orientation (SDO) and external motivation to respond without prejudice (EM), individuals more strongly evaluated the prejudicial beliefs they personally endorsed as not xenophobic. However, contrary to our hypotheses, individual levels of protestant work



ethic (PWE) did not substantially moderate the intra-individual association between prejudice endorsement and prejudice awareness. Also, the size of the moderating effect of SDO and HE decreased and became negligible, when simultaneously including interactive effects of IM, ES and EM in the analyses. This suggests that the (in)consistency of prejudice endorsement and prejudice awareness does not depend on general (anti-)egalitarian value orientations, but rather on their specific self-directed implications, as implied by self-related standards and motivations, such as IM, ES, and EM, that are particularly concerned with the personal endorsement of ethnic prejudice.

Overall, the findings from Study 1 corroborate a cognitive consistency approach to prejudice-related belief systems, suggesting that the subtlety of own ethnically prejudicial opinions indeed depends on (self-related) anti-prejudice standards. Individuals with high egalitarian standards, for whom this is relevant for the consistency of their belief system, seem to more strongly monitor the potentially prejudicial nature of their own opinions, in the first place, and endorse explicit prejudicial beliefs only to the extent that they are unaware of their xenophobic nature.

Due to the correlational nature of these data, we could, however, not unequivocally conclude that individual egalitarian values actually have a causal role in determining the intra-individual inconsistency, i.e., the negative association, of prejudice endorsement and prejudice awareness. In order to further examine this causal mechanism, we conducted a second study using an experimental research design. Additionally, since the moderating effects of (anti-)egalitarian standards found here are small, we sought to gain further confidence in their meaningfulness by replicating the current findings within this second study.

## Study 2

The goal of the second study is twofold. First, it aims to replicate the main findings from Study 1 based on data from an online survey. Second and most importantly, Study 2 seeks to extend the correlational findings from Study 1 by providing experimental evidence for the cognitive consistency perspective regarding the causal impact of egalitarian, prejudice-condemning values on the intra-individual association between prejudice endorsement and prejudice awareness. Specifically, we investigate in how far the salience of egalitarian standards triggers a stronger monitoring of own beliefs as potentially prejudiced as a means to

establish consistency with these salient standards, which would be mirrored by a stronger negative intra-individual association between prejudice endorsement and prejudice awareness. In this logic, individuals for whom egalitarian standards are made salient should—in order to be consistent with these standards—mostly endorse prejudicial beliefs only to the degree that they are unaware of their prejudicial content.

Adapting procedures that have been widely employed within previous studies, we experimentally manipulated the temporary salience of egalitarian standards in two ways. In one experimental condition, people were asked to describe failures to live up to egalitarian anti-prejudice standards in reflecting upon situations in which they exhibited ethnically prejudiced reactions (egalitarian standard failure condition; see e.g., Monteith, Mark, & Ashburn-Nardo, 2010; Moskowitz, 2002; Moskowitz & Li, 2011; Moskowitz, Li, Ignarri, & Stone, 2011). In the second experimental condition, people were asked to recall successful adherence to egalitarian standards, describing situations when they reacted in completely unprejudiced ways towards migrants (egalitarian standard affirmation condition; see e.g., Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011).

Informed by previous findings (e.g., Fehr & Sassenberg, 2010; Monteith et al., 2010; Monteith & Mark, 2005; Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011) and based on a cognitive consistency perspective (Festinger, 1957; as well as in line with related frameworks by Higgins, 1987; Wicklund & Gollwitzer, 1981, 1982), we reason that the egalitarian standard failure manipulation would—in the sense of an average treatment effect across participants—disturb participants' cognitive consistency in making people aware of a discrepancy between egalitarian standards and their described prejudiced reactions. The awareness of this discrepancy should be uncomfortable, as mirrored by increased negative affective reactions, and should trigger individual compensatory mechanisms. These compensatory responses should, in turn, be aimed at *restoring* cognitive consistency, i.e., the consistency with these salient egalitarian standards, and an unprejudiced self-image. We hypothesize that the salience of egalitarian standards would become manifest in an increased self-monitoring of own beliefs as potentially ethnically prejudiced and hence a stronger negative intra-individual association between prejudice endorsement and prejudice awareness for individuals in the egalitarian standard

failure compared to individuals in the control condition.<sup>11</sup>

Regarding the egalitarian standard affirmation condition, we have competing hypotheses. On the one hand, when regarding personal egalitarian standards (or an egalitarian self-image) in terms of a goal individuals pursue (e.g., Wicklund & Gollwitzer, 1981, 1982), one might argue that affirming people as unprejudiced can, in satisfying personal goals to be egalitarian or suggesting that one already lives up to a valued egalitarian self-image, shut down the further active pursuit of these egalitarian standards. Indeed, previous research has shown that such an affirmation of being egalitarian can subsequently undermine regulatory efforts to control stereotyping processes (e.g., on reaction time tasks; Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011), or have a so-called licensing effect leading to increased prejudiced responding on subsequent explicit prejudice measures (on which responses were ambiguous regarding the extent to which they were perceived as reflective of prejudice by the participants; Monin & Miller, 2001). From this perspective, one might assume that people in the egalitarian standard affirmation condition would feel so confident about themselves in already successfully fulfilling their personal egalitarian standards, that they would be even less likely than those in the egalitarian standard failure and even in the control condition to actively pursue these egalitarian standards. This would lead to a decreased self-monitoring of own opinions as potentially prejudiced and thus a less negative intra-individual association between prejudice endorsement and awareness (licensing hypothesis). In contrast, however, when considering that individuals seek to avoid inconsistencies between standards, values, and beliefs they hold as well as behaviors they exhibit (Festinger, 1957)—or aim to reduce discrepancies with valued self-images they seek to attain (Higgins, 1987; Wicklund & Gollwitzer, 1981, 1982)—one might predict that reminding people of past unprejudiced reactions would make individuals less likely to subsequently show explicit responses that would obviously contradict a previously affirmed egalitarian self-image and would thus potentially lead to cognitive dissonance (see Mullen & Monin, 2016, for a discussion of consistency and licensing effects). Thus, the salience of egalitarian standards induced by reflections on past unprejudiced behavior might cause a motivated response behavior aimed at *preserving* cognitive consistency, i.e., the consistency with these egalitarian standards, and at maintaining an unprejudiced self-image. From this perspective, we would

predict that people in the egalitarian standard affirmation condition might exhibit a stronger critical self-monitoring of personal beliefs as potentially ethnically prejudiced and thus a stronger negative intra-individual association between prejudice endorsement and prejudice awareness compared to the control condition (positive consistency hypothesis).<sup>11</sup>

## Method

### *Participants and procedure*

Data were collected as an online survey by a professional survey institute, which distributed the survey among participants from a German online access panel. This had the advantage of being able to draw a sample that is age- and gender-representative of the general population. Participation in the survey was entirely voluntary and participants could quit the interview at any time. Individuals who sign up for the online access panel collect points for participating in studies, which they can later exchange for vouchers, money, or donations.

In order to avoid suggesting a relationship between the self-report measures on individuals' prior egalitarian standards and the experimental manipulation of egalitarian standards salience, participants were told that they would participate in a study consisting of two separate parts, with the first assessing personal characteristics and aspects they consider important in life, and the second assessing opinions on different social and political topics such as migration. In the first part of the survey, egalitarian standards of the participants were assessed, i.e., internal (IM) and external motivation (EM) to respond without prejudice and egalitarian self-perception (ES) as the most important constructs from Study 1. Then, empathy was assessed with three subscales (12 items) of the German version of the Interpersonal Reactivity Index (Davis, 1983; German version: Paulus, 2009), which solely served as a neutral filler between the assessment of egalitarian standards and the experimental manipulation and is therefore not further reported on. Participants were then randomly assigned to one of three conditions: the two experimental conditions, i.e., the egalitarian standard failure condition and the egalitarian standard affirmation condition, and a neutral control condition. After that, they were first asked in how far they experienced different affective states (i.e., negative self-directed affect, discomfort and positive affect), which served as a manipulation check for the egalitarian standard failure condition (T1). Then, participants should indicate their agreement with

different prejudice items (prejudice endorsement) that varied in the strength of their prejudicial content. The items were similar to those used in Study 1, but the list of items was more extensive (see below). Next, socio-demographic characteristics were obtained. Subsequently, participants were informed that they would now be exposed to the same prejudice items they had evaluated beforehand, but should now give a different judgment, indicating in how far they evaluate these statements as xenophobic (prejudice awareness). Thereafter, the affect measure was administered for a second time (T2).<sup>12</sup>

The complete sample (after excluding one participant with only missing values on the outcome variable) comprised 1,201 participants (597 women, 604 men, 0 diverse/no indication;  $M_{age} = 44.62$  years,  $SD = 14.80$ , age range: 18–69 years), with subsamples of  $n = 450$  in the egalitarian standard failure,  $n = 259$  in the egalitarian standard affirmation, and  $n = 492$  in the control condition. Before running our analyses, we examined the content of the descriptions that participants provided in the three conditions to be able to identify those individuals who did not properly fulfill the tasks used for the experimental manipulations, which was essential for the successful induction of egalitarian standards salience. Although this has not been explicitly reported as a problem in previous studies using this manipulation of egalitarian standards with student samples (e.g., Monteith et al., 2010; Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011), our qualitative examination of the descriptions revealed that participants from this age- and gender-representative sample did not uniformly comply with the given instructions, especially in the two experimental conditions. We therefore decided to conduct our analyses and report the results both, on the full sample and on a ‘cleaner’ version of the sample, henceforth referred to as complete sample and adjusted sample, respectively. In order to avoid the exclusion of too many cases also within the adjusted sample, we decided to only remove the data from those individuals who provided an unequivocally improper answer or no description at all for the tasks given. Based on this general premise, we identified 313 participants (26.06%) and excluded their records in the adjusted sample (although not further discussed, a comparable percentage of data exclusions has, for instance, also been reported by Moskowitz & Li, 2011, Study 1). The detailed criteria for the data exclusions in the three conditions and the number of excluded cases are presented in Table A2 in the [supplemental material](#). The adjusted sample comprised

888 participants (456 women, 432 men, 0 diverse/no indication;  $M_{age} = 43.96$  years,  $SD = 14.95$ , age range: 18–69 years), with subsamples of  $n = 257$  in the egalitarian standard failure,  $n = 168$  in the egalitarian standard affirmation and  $n = 463$  in the control condition, respectively.

### **Experimental materials**

Adapting experimental procedures from previous research (Monteith et al., 2010; Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011), participants in the egalitarian standard failure condition were instructed to describe a situation from their past, in which they caught themselves showing a reaction (e.g., thoughts, feelings, or a behavior) towards migrants that was unjustly based on negative prejudice and thus contradicted values of tolerance and egalitarianism. If they could not remember a specific situation, participants were instructed to describe in general how they might behave in prejudiced ways towards migrants. Adapting the manipulation employed by Moskowitz and colleagues (Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011), the instructions in the egalitarian standard affirmation condition were constructed analogously, but here participants were asked to describe how they have reacted towards migrants in unprejudiced ways and have thus lived up to values of tolerance and egalitarianism. In the control condition, participants were asked to remember when they went grocery shopping and should describe how they handled the payment at the supermarket checkout. In order to increase the likelihood that participants would carefully read the instructions, they could only continue the survey after having spent one minute on the page with these experimental and control group tasks.

### **Measures**

#### **Level 1 variables (item ratings nested within respondents)**

**Prejudice endorsement.** Participants were asked to evaluate 15 prejudice items, all of them representing negative attitudes towards migrants in Germany. Amongst them were the same four prejudice items from the Subtle and Blatant Prejudice Scale (Pettigrew & Meertens, 1995) as employed in Study 1. The remaining items were either taken from existing scales or self-developed, covering different migration-related topics such as economy, culture and inner security. The complete list of items and their sources is

presented in Table A3 of the [supplemental material](#). As in Study 1, only items were selected for which the wording was logically sound for endorsement as well as awareness ratings. Besides these theoretical criteria, the selection of the additional items was also based on the average xenophobia ratings they received on a separate pretest.<sup>13</sup> We chose items that differed in the strength of their prejudicial content, thus covering a range from having been judged as rather not xenophobic to very xenophobic on that pretest. The main purpose of this item selection was to again ensure sufficient within-subject variation in item ratings, which was essential for this study. Participants indicated their agreement with the prejudice items on a 1 (*completely disagree*) to 5 (*completely agree*) scale. Mean endorsement ratings and standard deviations for these items ranged between  $M = 2.09$ ,  $SD = 1.05$  for the item with the lowest and  $M = 3.86$ ,  $SD = 1.04$  for the item with the highest average level of endorsement across participants in the complete sample (adjusted sample: between  $M = 1.93$ ,  $SD = 1.07$  and  $M = 3.82$ ,  $SD = 1.03$ , respectively).

**Prejudice awareness.** Afterwards, participants indicated for the same 15 prejudice items, in how far they evaluate these as xenophobic on a 1 (*not xenophobic at all*) to 5 (*very xenophobic*) scale. Mean awareness ratings and standard deviations for these items ranged between  $M = 2.34$ ,  $SD = 1.11$  for the item with the lowest and  $M = 3.81$ ,  $SD = 1.31$  for the item with the highest average awareness rating across participants in the complete sample (adjusted sample: between  $M = 2.35$ ,  $SD = 1.10$  and  $M = 4.00$ ,  $SD = 1.23$ , respectively).

As in Study 1, endorsement and awareness ratings for all prejudice items were treated as separate observations clustered within individuals in our analyses and were thus not averaged per respondent.

### Level 2 (individual difference) variables

**Internal and external motivation to respond without prejudice.** Participants' motivations to respond without prejudice were assessed with the full IM and EM scale (Plant & Devine, 1998), which were translated into German and rephrased to assess motivations to control prejudice towards migrants. The IM (complete sample:  $M = 3.54$ ,  $SD = 1.03$ ,  $\alpha = .92$ ; adjusted sample:  $M = 3.67$ ,  $SD = 0.99$ ,  $\alpha = .92$ ) and EM (complete sample:  $M = 2.61$ ,  $SD = 0.80$ ,  $\alpha = .72$ ; adjusted sample:  $M = 2.58$ ,  $SD = 0.78$ ,  $\alpha = .71$ ) items were averaged for each individual. Respondents indicated their

agreement with the respective items on a 1 (*completely disagree*) to 5 (*completely agree*) scale.

**Egalitarian self-perception.** Respondents' ES was measured with same item as in Study 1 (complete sample:  $M = 4.01$ ,  $SD = 1.02$ ; adjusted sample:  $M = 4.15$ ,  $SD = 0.93$ ). Respondents indicated their agreement on a 1 (*completely disagree*) to 5 (*completely agree*) scale.

**Affect.** As manipulation checks for the egalitarian standard failure condition we included measures of respondents' affect at two time points, i.e., immediately after the manipulation and control tasks (T1) and after the assessment of prejudice endorsement and prejudice awareness (T2). Adapting items from Monteith (1993, 1996), three different affective responses were measured with five items each: Negative self-directed affect (guilty, annoyed with myself, shameful, disappointed with myself, self-critical; T1: complete sample:  $M = 1.85$ ,  $SD = 0.79$ ,  $\alpha = .87$ ; adjusted sample:  $M = 1.82$ ,  $SD = 0.77$ ,  $\alpha = .86$ ), discomfort (bothered, uneasy, uncomfortable, tense, and anxious; T1: complete sample:  $M = 2.10$ ,  $SD = 0.87$ ,  $\alpha = .87$ ; adjusted sample:  $M = 2.04$ ,  $SD = 0.84$ ,  $\alpha = .87$ ) and positive affect (friendly, happy, optimistic, content, good; T1: complete sample:  $M = 3.26$ ,  $SD = 0.88$ ,  $\alpha = .89$ ; adjusted sample:  $M = 3.29$ ,  $SD = 0.89$ ,  $\alpha = .90$ ). Respondents rated on a 1 (*not at all*) to 5 (*very*) scale, in how far they experienced these affective states at the moment.

### Data structure and analytic strategy

Most importantly, also in Study 2, we aim to assess the intra-individual association of the endorsement of the different prejudice items and the evaluation of the same items as xenophobic (prejudice awareness), and again treated the 15 endorsement and awareness ratings for the prejudice items as paired repeated observations clustered within participants. Within this two-level hierarchical data structure, 18,015 paired endorsement and awareness ratings (level 1) were nested within 1,201 respondents (level 2) (adjusted sample: 13,320 awareness ratings were clustered within 888 respondents). First, we intended to replicate the findings from Study 1, by conducting the same multi-level regression analyses for participants in the control condition of Study 2. Second, we assessed the causal impact of the experimental induction of egalitarian standards salience on the intra-individual association between prejudice endorsement and prejudice awareness. To this end, we again estimated multilevel



**Table 3.** Replication of Study 1's multilevel regression analyses of the intra-individual effect of prejudice endorsement (level 1) and its cross-level interaction with individual difference characteristics (level 2) on prejudice awareness (control condition, complete sample).

Parameters and variables	Model 0 <i>b</i> (SE)	Model 1 <i>b</i> (SE)	Model 2 <i>b</i> (SE)	Model 3 <i>b</i> (SE)	Model 4 <i>b</i> (SE)	Model 5 <i>b</i> (SE)
<b>Fixed</b>						
Level 1						
Intercept	−.077 (.033)	.079 (.036)	.100 (.033)	.084 (.035)	.081 (.035)	.098 (.033)
Prejudice endorsement		−.357 (.018)	−.342 (.018)	−.347 (.018)	−.353 (.018)	−.338 (.017)
Level 2						
IM			.298 (.024)			.282 (.034)
ES				.217 (.026)		.007 (.035)
EM					.090 (.027)	.058 (.024)
<b>Cross-level interactions:</b>						
Prejudice endorsement x						
IM			−.070 (.016)			−.059 (.023)
ES				−.072 (.017)		−.029 (.023)
EM					.060 (.017)	.062 (.016)
<b>Random</b>						
Random intercept ( $\sigma_0^2$ )	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
	.515 (.035)	.332 (.024)	.250 (.018)	.290 (.021)	.322 (.023)	.246 (.018)
Random slope ( $\sigma_1^2$ )		.097 (.008)	.092 (.008)	.093 (.008)	.094 (.008)	.088 (.008)
Covariance ( $\sigma_{01}^2$ )		−.015 (.010)	.005 (.009)	−.001 (.009)	−.021 (.010)	.001 (.008)
Residual variance	.478 (.008)	.254 (.004)	.254 (.004)	.254 (.004)	.254 (.004)	.254 (.004)
ICC	.518					
BIC	16,924.980	13,159.780	13,030.500	13,101.580	13,151.580	13,044.290
Deviance	16,898.265	12,981.645	12,834.561	12,905.636	12,955.637	12,812.719

Based on  $n = 7,380$  item ratings clustered within  $n = 492$  participants. Predictor and outcome variables were z-standardized on their respective level of analysis and on the complete sample. Models 1–5 included a categorical control variable for the prejudice items. The BIC and deviance (−2 log-likelihood) statistics give an indication of model fit, with lower values representing a better fit.

regression models with prejudice awareness as the outcome and prejudice endorsement as the predictor variable (level 1), now including cross-level interactions with the condition, participants were assigned to, as a categorical predictor variable on level 2. Again, all multilevel models were estimated with maximum likelihood estimation and unrestricted covariance between the random effects using the xtmixed command in Stata (Version 15.1; StataCorp, 2017). For the assessment of the affective reactions in the three conditions, OLS regression models were estimated with affect measures (T1 and T2, respectively) as the outcome variables and condition as the categorical predictor variable. For all multilevel and OLS regression models reported here, the outcome and continuous predictor variables were z-standardized on their respective level of analysis and on the sample (i.e., complete or adjusted) to which is referred to, beforehand.

## Results

### Replication of findings from Study 1

In order to replicate the main findings from Study 1, we first ran the same multilevel regression models from Study 1 for participants in the control condition of the complete sample ( $n = 7,380$  item ratings clustered within  $n = 492$  participants) and the adjusted sample ( $n = 6,945$  item ratings clustered within  $n = 463$  participants). The simple intra-individual effect of prejudice endorsement on prejudice

awareness, which we first estimated based on a model without any control variables, was negative and of moderate size ( $b = -.475$ ,  $SE = .018$ ; adjusted sample:  $b = -.482$ ,  $SE = .017$ ; see Table 2). We then again included a categorical control variable for the specific prejudice items in the main models reported hereafter (Models 1–5; see Table 3 (complete sample) and Table 4 (adjusted sample)). Further corroborating findings from Study 1, the analyses revealed a negative main effect of prejudice endorsement on prejudice awareness (Model 1;  $b = -.357$ ,  $SE = .018$ ; adjusted sample:  $b = -.354$ ,  $SE = .018$ ), i.e., a moderate negative intra-individual association between prejudice endorsement and awareness: The more individuals endorsed a prejudicial belief, the less they thus evaluated the same belief as xenophobic. Models including cross-level interaction terms further corroborated the moderating role of individual (anti-)egalitarian standards and motivations on the intra-individual association between prejudice endorsement and prejudice awareness. An increase in the endorsement of a prejudice item was more strongly related to a decrease in prejudice awareness for the same item for individuals with higher levels of IM (Model 2;  $b = -.070$ ,  $SE = .016$ ; adjusted sample:  $b = -.059$ ,  $SE = .015$ ) and ES (Model 3;  $b = -.072$ ,  $SE = .017$ ; adjusted sample:  $b = -.055$ ,  $SE = .015$ ). Moreover, endorsement and awareness ratings were intra-individually less negatively associated and more unrelated with increasing levels of EM (Model 4;  $b = .060$ ,  $SE = .017$ ; adjusted sample:  $b = .055$ ,

**Table 4.** Replication of Study 1's multilevel regression analyses of the intra-individual effect of prejudice endorsement (level 1) and its cross-level interaction with individual difference characteristics (level 2) on prejudice awareness (control condition, adjusted sample).

Parameters and variables	Model 0 <i>b</i> ( <i>SE</i> )	Model 1 <i>b</i> ( <i>SE</i> )	Model 2 <i>b</i> ( <i>SE</i> )	Model 3 <i>b</i> ( <i>SE</i> )	Model 4 <i>b</i> ( <i>SE</i> )	Model 5 <i>b</i> ( <i>SE</i> )
<b>Fixed</b>						
Level 1						
Intercept	.131 (.034)	.061 (.037)	.113 (.034)	.088 (.035)	.062 (.036)	.111 (.034)
Prejudice endorsement		-.354 (.018)	-.345 (.018)	-.350 (.018)	-.352 (.017)	-.344 (.017)
Level 2						
IM			.294 (.024)			.267 (.033)
ES				.215 (.025)		.025 (.033)
EM					.090 (.028)	.063 (.024)
<b>Cross-level interactions:</b>						
Prejudice endorsement x						
IM			-.059 (.015)			-.056 (.021)
ES				-.055 (.015)		-.015 (.021)
EM					.055 (.016)	.056 (.016)
<b>Random</b>						
Random intercept ( $\sigma_0^2$ )	Estimate .505 (.035)	Estimate .327 (.024)	Estimate .242 (.018)	Estimate .281 (.021)	Estimate .316 (.024)	Estimate .237 (.018)
Random slope ( $\sigma_1^2$ )		.088 (.008)	.084 (.008)	.086 (.008)	.086 (.008)	.081 (.008)
Covariance ( $\sigma_{01}^2$ )		-.023 (.010)	-.006 (.008)	-.011 (.009)	-.029 (.010)	-.010 (.008)
Residual variance	.491 (.009)	.259 (.005)	.258 (.005)	.258 (.005)	.259 (.005)	.258 (.005)
ICC	.507					
BIC	16,095.600	12,475.550	12,351.250	12,418.770	12,467.270	12,365.470
Deviance	16,069.062	12,298.639	12,156.645	12,224.164	12,272.667	12,135.476

Based on  $n = 6,945$  item ratings clustered within  $n = 463$  participants. Predictor and outcome variables were z-standardized on their respective level of analysis and on the adjusted sample. Models 1–5 included a categorical control variable for the prejudice items. The BIC and deviance ( $-2$  log-likelihood) statistics give an indication of model fit, with lower values representing a better fit.

$SE = .016$ ). Within the final model (Model 5), including all three interaction terms, the interactive effects of IM and EM with prejudice endorsement remained robust ( $b = -.059$ ,  $SE = .023$  (adjusted sample:  $b = -.056$ ,  $SE = .021$ ); and  $b = .062$ ,  $SE = .016$  (adjusted sample:  $b = .056$ ,  $SE = .016$ ), respectively), whereas the size of the interactive effect of prejudice endorsement and ES decreased ( $b = -.029$ ,  $SE = .023$ ; adjusted sample:  $b = -.015$ ,  $SE = .021$ ). Compared to the first model including only prejudice endorsement as a (level 1) predictor variable for prejudice awareness (BIC = 13,159.780; adjusted sample: BIC = 12,475.550), the full model exhibited and increased fit with the data (BIC = 13,044.290; adjusted sample: BIC = 12,365.470). Additionally, the inclusion of the individual difference (level 2) variables and their cross-level interaction effects with prejudice endorsement (level 1) contributed to the explanation of random slope variance, which was reduced by 9.28% (adjusted sample: 7.95%) from Model 1 to Model 5.<sup>14</sup>

## Experimental results

### Manipulation checks

Results from OLS regression analyses (see Table A6 in the supplemental material) reveal that the egalitarian standard failure manipulation—compared to the control condition—increased negative self-directed affect ( $b = .389$ ,  $SE = .064$ ; adjusted sample:  $b = .613$ ,  $SE = .075$ ) as well as feelings of discomfort ( $b = .235$ ,

$SE = .065$ ; adjusted sample:  $b = .323$ ,  $SE = .077$ ), and decreased positive affect ( $b = -.346$ ,  $SE = .064$ ; adjusted sample:  $b = -.444$ ,  $SE = .076$ ) immediately after the experimental manipulation (T1). Only small differences in affect were observed between participants in the egalitarian standard affirmation and the control condition ( $bs$  between  $-.048$  and  $.153$ ; adjusted sample:  $bs$  between  $.009$  and  $.089$ ). This indicates that the egalitarian standard failure manipulation indeed triggered an emotionally aversive state of cognitive dissonance. The same OLS regression models with affect at T2, i.e., after participants gave their prejudice endorsement and prejudice awareness ratings, showed a reduction of cognitive dissonance for participants in the egalitarian standard failure condition. Although they still exhibited somewhat increased negative self-directed affect ( $b = .110$ ,  $SE = .065$ ; adjusted sample:  $b = .242$ ,  $SE = .077$ ) and lower positive affect ( $b = -.179$ ,  $SE = .065$ ; adjusted sample:  $b = -.232$ ,  $SE = .077$ ) at T2 than participants in the control condition, these differences were yet substantially smaller than at T1. Differences in experienced discomfort between participants in the egalitarian standard failure condition and the control group almost vanished at T2 ( $b = .036$ ,  $SE = .065$ ; adjusted sample:  $b = .079$ ,  $SE = .078$ ). This indicates that participants' responses between T1 and T2 alleviated the aversive affective consequences of the egalitarian standard failure manipulation task. As at T1, only small differences in affect were observed at T2

**Table 5.** Multilevel regression analyses of the effect of egalitarian standards salience on the intra-individual association between prejudice endorsement and prejudice awareness.

Parameters and variables	Complete sample			Adjusted sample		
	Model 0 <i>b</i> (SE)	Model 1 <i>b</i> (SE)	Model 2 <i>b</i> (SE)	Model 0 <i>b</i> (SE)	Model 1 <i>b</i> (SE)	Model 2 <i>b</i> (SE)
<b>Fixed</b>						
Level 1						
Intercept	.000 (.021)	.154 (.023)	.107 (.030)	.000 (.023)	.179 (.026)	.098 (.031)
Prejudice endorsement		−.373 (.011)	−.357 (.017)		−.389 (.012)	−.348 (.016)
Level 2						
Condition <sup>a</sup>						
Egalitarian standard failure			.068 (.037)			.169 (.043)
Egalitarian standard affirmation			.101 (.044)			.170 (.049)
Cross-level interactions:						
Prejudice endorsement ×						
Condition <sup>a</sup>						
Egalitarian standard failure			−.032 (.023)			−.082 (.026)
Egalitarian standard affirmation			−.015 (.027)			−.078 (.030)
<b>Random</b>						
Random intercept ( $\sigma_0^2$ )	Estimate .479 (.021)	Estimate .298 (.014)	Estimate .296 (.014)	Estimate .449 (.023)	Estimate .276 (.015)	Estimate .270 (.015)
Random slope ( $\sigma_1^2$ )		.084 (.005)	.084 (.005)		.073 (.005)	.072 (.005)
Covariance ( $\sigma_{01}^2$ )		−.006 (.006)	−.006 (.006)		−.003 (.006)	.000 (.006)
Residual variance	.521 (.006)	.288 (.003)	.288 (.003)	.551 (.007)	.290 (.004)	.290 (.004)
ICC	.479			.449		
BIC	42,644.120	33,714.790	33,746.100	32,184.330	24,927.380	24,931.370
Deviance	42,614.720	33,518.808	33,510.924	32,155.838	24,737.442	24,703.444

Based on  $N = 18,015$  item ratings clustered within  $N = 1,201$  participants (complete sample) or  $N = 13,320$  item ratings clustered within  $N = 888$  participants (adjusted sample). Continuous predictor and outcome variables were z-standardized on their respective level of analysis and on the respective sample (i.e., complete or adjusted). Condition was entered as an unstandardized categorical variable. Models 1–2 included a categorical control variable for the prejudice items. The BIC and deviance ( $-2$  log-likelihood) statistics give an indication of model fit, with lower values representing a better fit.

<sup>a</sup>Reference category: Control condition.

between participants in the egalitarian standard affirmation and the control condition (*bs* between  $-.063$  and  $.149$ ; adjusted sample: *bs* between  $-.041$  and  $.114$ ).

### Impact of egalitarian standards salience

We initially ran the analyses to examine the impact of the egalitarian standard manipulation based on the complete, unadjusted sample. Multilevel models were calculated with prejudice awareness as the outcome variable and the endorsement for the prejudice items as the predictor variable (level 1). A model containing no control variables overall revealed a negative ‘raw’ intra-individual effect of prejudice endorsement on prejudice awareness ( $b = -.493$ ,  $SE = .011$ ; see Table 2, left panel). We then conducted the main analyses, including a categorical control variable for the prejudice items in the models (Models 1–2; see Table 5, left panel). Model 1 revealed a negative main effect of prejudice endorsement on prejudice awareness of moderate strength ( $b = -.373$ ,  $SE = .011$ ), i.e., also across conditions there was a negative intra-individual association between prejudice endorsement and awareness. In order to assess the potential moderating effect of the condition on this association between prejudice endorsement and awareness, we then calculated multilevel regression models with prejudice awareness as the outcome variable, and prejudice endorsement (level 1) and condition (categorical level 2 variable) as well as their cross-level interaction term as

predictor variables (Model 2). This model, based on the complete, unadjusted sample, revealed very small moderating effects of egalitarian standards salience on the intra-individual association between prejudice endorsement and prejudice awareness. Participants in the egalitarian standard failure condition exhibited a slightly stronger negative intra-individual association between prejudice endorsement and prejudice awareness ( $b = -.032$ ,  $SE = .023$ ) compared to participants in the control condition. The same is true for individuals assigned to the egalitarian standard affirmation condition ( $b = -.015$ ,  $SE = .027$ ; see also Table 2, left panel, for an overview on the simple intra-individual effects of prejudice endorsement on prejudice awareness, estimated based on separate multilevel regression models for the three conditions containing no control variables). These results might be interpreted as first, very tentative support for our hypotheses, providing a rather conservative estimation of the effect of the experimental conditions, since participants did not uniformly fulfill the tasks employed for the experimental manipulation of egalitarian standards salience.

We therefore conducted, in a second step, the same analyses on the impact of the experimental manipulation based on the adjusted sample, in which individuals who unequivocally disregarded the manipulation instructions were excluded. The simple intra-individual effect of prejudice endorsement on prejudice awareness estimated based on a model without

control variables was again negative and of moderate strength ( $b = -.535$ ,  $SE = .012$ ; see Table 2, right panel). The multilevel regression models for the main analyses then included a categorical control variable for the prejudice items (Models 1–2; see Table 5, right panel). Model 1 also revealed a moderate negative main effect of prejudice endorsement on prejudice awareness ( $b = -.389$ ,  $SE = .012$ ), i.e., a negative intra-individual association between prejudice endorsement and awareness across conditions. Model 2 with prejudice awareness as the outcome variable, and prejudice endorsement (level 1) and condition (categorical level 2 variable) as well as their cross-level interaction term as predictor variables, showed that participants in the egalitarian standard failure condition exhibited a stronger negative intra-individual association between prejudice endorsement and prejudice awareness ( $b = -.082$ ,  $SE = .026$ ) compared to participants in the control condition. The same is true for individuals assigned to the egalitarian standard affirmation condition ( $b = -.078$ ,  $SE = .030$ ; see also Table 2, right panel, for the simple intra-individual effects of prejudice endorsement on prejudice awareness, estimated based on separate multilevel regression models for the three conditions containing no control variables). Thus, the size of the moderating effects of egalitarian standards salience on the intra-individual association between prejudice endorsement and prejudice awareness increased within these models based on the adjusted sample compared to the models estimated based on the complete, unadjusted sample.

## Discussion

First, this second study was able to replicate findings from Study 1—for participants in the control condition—based on data from an online survey. Again, we overall found a negative intra-individual association between prejudice endorsement and prejudice awareness, which was of moderate strength and comparable to the effect size found in Study 1. Also, results from Study 2 could further corroborate a cognitive consistency perspective on the mechanisms of subtle ethnic prejudice. In line with results from Study 1, we found that with higher self-related egalitarian standards and motivations (i.e., IM and ES) and lower anti-egalitarian standards (i.e., EM), the negative intra-individual association between prejudice endorsement and prejudice awareness became more pronounced. Thus, as this is relevant for the consistency of their belief system, individuals with higher standards that condemn and lower standards that condone personal

endorsement of ethnic prejudice seemed to more strongly monitor their own beliefs as potentially prejudiced and were, in turn, more unaware of the xenophobic content (or more convinced of the unprejudicial nature) of the prejudicial beliefs they did personally endorse. These small interactive effects were also of comparable size for the two studies. The only difference in the findings of the two studies is that—unlike in Study 1—the size of the moderating effect of ES decreased in Study 2 when all interactive effects (i.e., IM, EM and ES with prejudice endorsement, respectively) were included simultaneously in the model.

Most importantly and extending these correlational findings, Study 2 could provide experimental evidence on the causal role of egalitarian standards in shaping the intra-individual association between prejudice endorsement and prejudice awareness. In line with our prediction, participants in the egalitarian standard failure condition exhibited a slightly stronger negative intra-individual association between prejudice endorsement and prejudice awareness compared to individuals in the control condition. In this condition the affect measures allowed us to trace the hypothesized cognitive consistency mechanism behind the egalitarian standards manipulation more narrowly: Indeed, thinking about past prejudiced reactions initially caused cognitive dissonance (most importantly reflected by increased levels of negative self-directed affect), which was however alleviated after participants had given their prejudice endorsement and awareness ratings. This indicates that participants' endorsement and awareness ratings served as a compensatory response to live up to egalitarian standards and to restore cognitive consistency, i.e., the consistency with these salient egalitarian standards. Furthermore, individuals in the egalitarian standard affirmation condition overall exhibited a stronger negative intra-individual association between prejudice endorsement and prejudice awareness compared to control group participants. This effect was of similar strength as the effect of the egalitarian standard failure manipulation. Thus, in line with our positive consistency hypothesis, thinking about past instances of adherence to egalitarian standards indeed led to an overall response pattern of prejudice endorsement and awareness that would further preserve the consistency between the explicit evaluations of migration-related statements and these egalitarian standards. The effects presented here should thereby be regarded as average effects of the experimental treatments across participants, with the cognitive consistency mechanism hypothesized to



underlie the induction of egalitarian standards salience in the two experimental conditions being mainly applicable to individuals with higher prior egalitarian standards.<sup>11</sup>

Overall, the size of the effect of both egalitarian standard manipulations should be interpreted with some caution. Due to the observed difficulties for participants in this population sample to fulfill the tasks employed for the experimental manipulations, we chose to present both, the effects of the experimental conditions estimated based on the complete sample—as rather attenuated effects, since participants who did not fulfill the experimental task were included—and the effects estimated based on an adjusted sample, excluding participants who unequivocally disobeyed the instructions—as stronger effects, whereby however a bias due to the selectivity of the compliers cannot be completely ruled out. The size of the moderating effect of the experimental manipulations on the intra-individual association between prejudice endorsement and awareness estimated based on the adjusted sample is comparable to (or even a bit larger than) the size of the moderating effects of a 1 SD-increase or 1 SD-decrease in individual egalitarian (i.e., IM, and ES) or anti-egalitarian (i.e., EM) standards, respectively, as indicated by the correlational findings in Study 1 and Study 2.

Despite these qualifications concerning the interpretation of the effect sizes, the experimental findings from Study 2 provide important first evidence on the causal role of egalitarian standards in the cognitive mechanism behind subtle ethnic prejudice. Results could reveal that participants for whom egalitarian standards were made salient—by reflecting upon instances of past prejudiced or unprejudiced reactions—exhibited a stronger negative intra-individual association between prejudice endorsement and prejudice awareness. This suggests that the temporary salience of egalitarian, prejudice-condemning standards indeed triggered or enhanced a self-monitoring of own opinions as potentially prejudiced. Individuals for whom anti-prejudice standards were made salient devoted more vigilance to their evaluation of migration-related statements and endorsed prejudicial beliefs only to the extent that they are unaware of their xenophobic content, thereby establishing, i.e., restoring or preserving, consistency with these egalitarian standards.

## General discussion

Since in today's societies the endorsement of ethnic prejudice is generally condemned by prevailing

egalitarian anti-prejudice norms, it seems intuitive to assume and has been implied by previous social-psychological research, that many individuals who explicitly endorse ethnically prejudicial beliefs are largely unaware of their xenophobic content, based on the presumption that people would find it nowadays rather aversive to perceive themselves (and also to be perceived by others) as prejudiced. In that vein, it has thus been suggested that one's own ethnic prejudice is likely to be subtle to oneself and that making individuals aware of their own ethnically prejudicial beliefs is therefore an effective means to motivate them to reconsider their opinions. However, earlier studies have actually not directly investigated how people evaluate their own prejudicial opinions in the first place, and whether the explicit endorsement of an ethnically prejudicial statement (prejudice endorsement) and the simultaneous explicit evaluation of this belief as xenophobic or racist (prejudice awareness) are really rather inconsistent with each other.

Two studies addressed this issue, in examining the intra-individual association between prejudice endorsement and prejudice awareness, i.e., in how far individuals do or do not judge ethnically prejudicial statements they personally endorse as xenophobic. Drawing on a cognitive consistency perspective (e.g., Festinger, 1957; Gawronski, 2012), we also investigated in how far the association of prejudice endorsement and prejudice awareness might depend on individual standards and motivations condemning or condoning the personal endorsement of ethnic prejudice (Study 1 and Study 2) or on the salience of these standards (Study 2).

In line with the basic rationale underlying approaches on contemporary forms of explicit ethnic prejudice (e.g., McConahay, 1986; Pettigrew & Meertens, 1995), our analyses in both studies, based on data from a telephone and an online survey, overall reveal an inconsistency between prejudice endorsement and prejudice awareness as indicated by their negative intra-individual association: The more individuals agreed with an ethnically prejudicial statement, the less they tended to evaluate this statement as xenophobic. However, this negative intra-individual relationship was only of moderate strength, and not as strong as suggested by previous research that has indirectly approached this question (e.g., Crandall et al., 2002). Thus, our findings also indicate that intra-individually, prejudice endorsement and prejudice awareness, despite being negatively related, are not perfect mirror images of each other. Hence, one can neither straightforwardly infer from the

endorsement of an ethnically prejudicial statement (e.g., in a survey) that individuals are unaware of its xenophobic content, nor that people who reject such a statement are necessarily aware that it represents ethnic prejudice.

Moreover, the correlational findings from both studies reveal that this negative intra-individual association between prejudice endorsement and prejudice awareness was more pronounced for individuals with higher egalitarian and lower anti-egalitarian standards. Results also indicated that especially self-related anti-prejudice standards such as internal motivation to respond without prejudice (IM) egalitarian self-perception (ES) and external motivation to respond without prejudice (EM), rather than general (anti-)egalitarian value orientations, are important in shaping that association and thus the (in)consistency of prejudice endorsement and prejudice awareness. Thus, with higher self-related standards that condemn (IM, ES) and lower standards that condone (EM) the private endorsement of ethnic prejudice—which might also be conceptualized as self-determined standards to avoid personal prejudice endorsement (Legault et al., 2007)—, individuals more strongly evaluated the prejudicial beliefs they personally endorsed as not xenophobic, being more unaware of the prejudicial content of their own beliefs. These findings are congruent with our hypotheses we developed from a cognitive consistency perspective (e.g., Festinger, 1957; Gawronski, 2012), in acknowledging that being or perceiving oneself as unprejudiced is not equally important to all individuals. For individuals with higher standards condemning (and lower standards condoning) personal prejudice endorsement, on the one hand, the endorsement of a belief and its simultaneous evaluation as xenophobic are logically inconsistent, in colliding with their personal egalitarian values. In assuming that people would not endure the aversive emotional state of cognitive dissonance resulting from this inconsistency, it makes sense that especially these individuals are initially more strongly convinced that the prejudicial beliefs they endorse do not represent ethnic prejudice, exhibiting a stronger negative intra-individual association between prejudice endorsement and prejudice awareness. In other words, as it is relevant for the consistency of their belief system, these individuals are more likely to engage in a critical self-monitoring of own beliefs as potentially prejudiced, so that they endorse a prejudicial statement only to the extent that they are unaware of its prejudicial content (or evaluate a prejudicial statement they endorse as not xenophobic). On the other hand, for individuals with lower egalitarian (or higher anti-

egalitarian) standards, the evaluation of own beliefs as xenophobic should not collide with their personal values and should not cause cognitive inconsistency to the same degree. These individuals are less likely to monitor whether the beliefs they privately endorse are xenophobic, as it is mirrored by the weaker negative intra-individual association of prejudice endorsement and prejudice awareness. These findings are, in general, also congruent with basic assumptions from other frameworks related to cognitive dissonance theory, such as self-discrepancy theory (Higgins, 1987) or symbolic self-completion theory (Wicklund & Gollwitzer, 1981, 1982). From these perspectives, especially individuals for whom an 'unprejudiced self' represents an important, aspired self-definition or a personally relevant state of the self they strive for would seek to reduce discrepancies with this valued egalitarian self-representation, as it would be empirically reflected in our studies by a stronger negative association between prejudice endorsement and prejudice awareness.

The second study extended the correlational findings on the moderating impact of self-related egalitarian standards, in examining their causal role in initiating a self-monitoring process that entails a simultaneous consideration of prejudice endorsement and prejudice awareness based on an experimental research design. Indeed, our findings showed that the induction of egalitarian standards salience, by asking people to recall instances in which they behaved in ethnically prejudiced (egalitarian standard failure condition) or unprejudiced (egalitarian standard affirmation condition) ways, led to a stronger negative intra-individual association of prejudice endorsement and prejudice awareness. Thus, the salience of egalitarian standards triggered a critical reflection of own beliefs as potentially prejudiced as a cognitive process that served to restore (egalitarian standard failure condition) or preserve (egalitarian standard affirmation condition) consistency with these activated standards. In turn, individuals for whom egalitarian standards were made salient were more unaware of the xenophobic content (or more convinced of the unprejudicial nature) of those prejudicial beliefs they did personally endorse.

Although not the main purpose of our study, this evidence on the impact of the egalitarian standard affirmation condition can also contribute to previous research that has brought forward conflicting findings on consistency and licensing effects of previous moral behavior (see Mullen & Monin, 2016, for a review). Based on earlier research findings (Monin & Miller, 2001; Moskowitz, 2002; Moskowitz et al., 2011;

Moskowitz & Li, 2011) and when regarding personal egalitarian standards in terms of goals people pursue (e.g., Wicklund & Gollwitzer, 1981, 1982), one could have predicted that affirming people in already having achieved the goal of being unprejudiced would subsequently shut down the pursuit of egalitarian standards and the further monitoring of prejudiced responses. In contrast, however, and in line with the perspective that people would seek to avoid inconsistencies between standards they hold and beliefs they express (Festinger, 1957) or discrepancies with valued self-images (Higgins, 1987; Wicklund & Gollwitzer, 1981, 1982), our findings reveal that the recall of the successful adherence to egalitarian standards afterwards led, in the sense of an average effect of the manipulation across participants, to an increased critical monitoring of potentially prejudicial responses (i.e., positive consistency)—at least when these responses are explicit in nature and might cause an unequivocal inconsistency with the previously affirmed, salient egalitarian standards or the affirmed unprejudiced self-image (see e.g., also Monin & Miller, 2001, for this discussion).

Overall, the correlational and experimental findings from our two studies on the moderating role of personal egalitarian anti-prejudice standards (or the salience of such standards) in determining the strength of the negative intra-individual association of prejudice endorsement and prejudice awareness strongly corroborate a cognitive consistency approach to prejudice-related belief systems. Extending previous research in directly investigating how people evaluate their own ethnically prejudicial beliefs in the first place and how this is determined by egalitarian standards, the current study has, from a practical perspective, important implications for real-world prejudice-reduction interventions. Our findings from both studies demonstrate that those parts of anti-prejudice interventions (e.g., civic education campaigns) that aim to combat explicit ethnic prejudice by making people aware of their own prejudicial beliefs might be best suited for individuals with self-standards and motivations that condemn the private endorsement of (ethnic) prejudice, since they seem to routinely monitor whether their own beliefs are potentially prejudicial, and are initially, i.e., privately and prior to potential interventions, more strongly unaware of the xenophobic nature of the beliefs they personally endorse. This rationale is also in line with research findings by Plant and Devine (2009), who showed that individuals with primarily internal standards to be unprejudiced are, by default, not that interested in participating in prejudice

reduction programs, presumably convinced they do not need such an assistance, yet are even more motivated to take part in such programs, when made aware of their own prejudice. In the light of our results, we would argue that highly egalitarian individuals might initially be in an ‘equilibrium state’ of cognitive consistency, convinced that the explicit beliefs they personally hold are not xenophobic, which might to some degree also reflect a state of self-contentment. Yet, it is exactly this equilibrium state that, when being disturbed, provides a promising avenue for anti-prejudice approaches. For individuals with lower egalitarian standards, for whom prejudice endorsement and prejudice awareness are not that inconsistent and who are not that unaware of the xenophobic nature of the ethnically prejudicial beliefs they personally endorse beforehand, it might be a better idea to first strengthen exactly these self-standards that condemn personal endorsement of ethnic prejudice, which routinely initiate a critical self-monitoring of personal beliefs as potentially prejudiced.

From a theoretical and methodological perspective, the current study also emphasizes the need for a differentiated view on the subtlety of explicit ethnic prejudice. When aiming to target ‘subtlety’ on the individual level, in the sense of a person’s unawareness that certain beliefs are reflective of ethnic prejudice, our findings show that this has to be assessed directly and cannot simply be inferred from an individual’s endorsement of certain opinions. Also, our results on between-respondents differences in the intra-individual association of prejudice endorsement and prejudice awareness clearly reveal the necessity to approach within-person dynamics in the subtlety of (ethnic) prejudice in a direct manner (see e.g., Molenaar, 2004; Molenaar & Campbell, 2009, for this methodological discussion). The current study shows that the relationship between prejudice endorsement and prejudice awareness is more complex on an intra-personal level than one might indirectly conclude from research operating at a higher or different level of aggregation (e.g., Crandall et al., 2002; Swim et al., 2005).

### **Limitations and future directions**

Although providing important insights into cognitive mechanisms behind the subtlety of ethnic prejudice, this study nevertheless carries some limitations that offer avenues for further research.

To begin with, it needs to be noted that we adopted a simplistic view on cognitive (in)consistency

here, for the sake of clarity, and that prejudice-related belief systems are certainly in reality far more complex, comprising a multitude of interrelated propositions. We do neither intend to imply that propositions deriving from personal egalitarian standards, as the focus of the current study, are always uniformly accessible to individuals across different situations, nor that they are the only relevant moderating variables, and suggest that future research could also examine other personal standards that potentially determine the inconsistency (or even the consistency) between prejudice endorsement and awareness, for example individual motivations to (intentionally) express prejudice (Forscher, Cox, Graetz, & Devine, 2015). Additionally, we emphasize that the cognitive consistency perspective adopted here applies to private contexts, i.e., how individuals evaluate different beliefs on their own. We suppose that the degree to which individuals would publicly express opinions they endorse or would openly evaluate them as xenophobic, is of course subject to other processes going beyond this private logic of cognitive consistency. In public contexts, the individual sensitivity to external influences (e.g., social anti-prejudice norms) are likely to play a role. We would, for instance, expect that high-EM individuals might—in striving to keep up an unprejudiced self-image in front of others—respond differently in a public compared to the current, more private survey setting, in monitoring their own beliefs more strongly as potentially prejudicial or at least labeling their own beliefs more strongly as not xenophobic when their responses are observable by others. Also, it needs to be emphasized that the logic of cognitive consistency pertains to propositional processing (e.g., Gawronski, 2012) and that the perspective employed in this research thus specifically applies to explicit prejudicial beliefs, as opinions to which individuals consciously assign a truth value by endorsing or rejecting them. Beyond the scope of the current studies it might, however, be an interesting avenue for further investigations to extend the limited body of research that targets from a cognitive consistency perspective, how people integrate implicit ethnically prejudicial biases of which they might become aware (e.g., automatic stereotypical thoughts or negative affective or behavioral reactions) as propositional elements they reject or endorse into their belief systems and how this might depend on their personal egalitarian standards and motivations (see e.g., Gawronski & Bodenhausen, 2014; Gawronski et al., 2008, for this discussion).

Also, we cannot ultimately disentangle within our study in how far, especially for individuals with highly egalitarian self-standards, the evaluation of the prejudicial beliefs as not xenophobic represents ‘true’ or ‘naïve’ unawareness of their prejudicial nature or also involves a strategic component, with individuals deliberately labeling their own opinions as not xenophobic in order to prevent cognitive dissonance. We would tentatively suggest that both is true to some degree: Thus, we would assume that individuals who are genuinely motivated to not be (ethnically) prejudiced usually reject opinions they identify as xenophobic, and are thus unaware of the prejudicial content of those beliefs they do endorse and which have slipped under their radar. But we also suspect that, at the same time, these individuals might, if they do endorse prejudicial beliefs and (maybe hesitantly) identify them as such, also self-servingly define these beliefs as not xenophobic (even if they might realize that most others would) in order to avoid cognitive inconsistency. However, further research is warranted to investigate these two possible—self-critical and self-presentational—mechanisms more in-depth. Relating to this point, the current data do not allow us to draw any causal inferences on whether individuals adapt their endorsement of a prejudicial statement based on whether they evaluate it as xenophobic or vice versa. Indeed, our conclusions based on the cognitive consistency perspective employed within this study are independent of the direction of causation between prejudice endorsement and prejudice awareness and we would assume that the two judgments mutually influence each other, with especially highly egalitarian individuals indicating their endorsement with a statement based on the concurrent evaluation of whether it is xenophobic and vice versa. Nevertheless, it might be interesting for future research to consider, whether individuals endorse ethnically prejudicial beliefs mainly because they are unaware that they are xenophobic, or do not label their beliefs as xenophobic, because they endorse them.

Additionally, since it was the main purpose of the second study to establish the causal impact of egalitarian standards salience on the intra-individual association between prejudice endorsement and prejudice awareness, we analyzed and presented the effects of our manipulations as average treatment effects across participants. However, it should be noted that the hypothesized consistency mechanisms behind the manipulations should be primarily applicable to individuals with high prior egalitarian standards. Since such an analysis was beyond the scope of the current



article, it would be worthwhile for future studies to examine heterogeneous effects of the egalitarian standards manipulations employed here.

Moreover, there are some methodological aspects of the current study that need to be critically reflected on. First, the moderating effects of individual egalitarian standards found in Study 1 and Study 2 and of the experimentally induced egalitarian standards salience (Study 2) are small. On the one hand, we consider these small effects meaningful and also plausible in the context of the current study, as we expected the intra-individual association between prejudice endorsement and prejudice awareness to vary within a restricted range. Thus, we predicted that the relationship would inter-individually vary between a strongly negative association and no association at all, but did not suppose that people with low egalitarian standards (or for whom egalitarian standards were not made salient) would, in fact, show a positive relationship between endorsement and awareness. On the other hand, we however also think that the small effect sizes might be attributed to the fact that egalitarian standards can only unfold their moderating impact for people who are willing to engage in a self-monitoring of their beliefs—an individual readiness that might not be taken for granted. Future research might, for instance, examine the role of individual differences in need for cognition (Cacioppo & Petty, 1982) to investigate this potential mechanism further. Second, and relating to this latter consideration, it needs to be emphasized that the size of the effect of the experimental manipulations in Study 2 should be interpreted with some caution, since the qualitative examination of the material revealed that participants did not uniformly comply with the self-reflection instructions for the experimental tasks. Since such problems to fulfill the task instructions did not arise to the same degree within the control condition, we think that the failure to describe instances of having reacted in prejudiced (and also unprejudiced) ways cannot be attributed to a general lethargy on behalf of the online survey participants. It might rather reflect a lack of readiness (or maybe also the inability) to engage in critical self-monitoring of one's own attitudes and behaviors, especially regarding sensitive issues such as (ethnic) prejudice. In fact, it follows from the very cognitive consistency perspective employed here itself that especially asking people to recall instances in which they failed to live up to egalitarian standards (i.e., the exercise used for the egalitarian standard failure condition) might be a rather challenging task, as it requires at least individuals who

strongly value egalitarian ideals to voluntarily lead themselves into cognitive inconsistency and cognitive dissonance, as an aversive state people mostly seek to avoid. In the light of these considerations, future experimental research using different experimental designs, such as fake feedback manipulations or priming methods, into the effect of egalitarian standards salience is definitely needed. Despite these cautionary notes, we are nevertheless confident that the findings from our experimental study provide first important evidence on the causal role of egalitarian standards in initiating a self-monitoring and balancing process between prejudice endorsement and prejudice awareness. Additionally, we think that the observed unwillingness in parts of the sample to comply especially with the experimental task instructions carry important methodological as well as theoretical implications. First, they show that these widely used procedures (Monteith et al., 2010; Moskowitz, 2002; Moskowitz et al., 2011; Moskowitz & Li, 2011) to induce egalitarian standards salience might not be readily applicable beyond college samples. This stresses the importance of conducting also experimental research with larger samples that are more representative of the general population. Second, the very fact that many people especially in the egalitarian standard failure condition seem to be reluctant to critically self-reflect their opinions as potentially prejudiced is an important theoretical finding in itself. It shows that engaging in self-criticism and especially reflecting on violations of own values, which is at the same time a crucial step for the self-regulation of prejudicial responses and successful adherence to egalitarian standards, is—in line with basic assumptions of consistency frameworks—necessarily a demanding endeavor. It could also indicate that cognitive consistency processes in general, including that of interest in the current article, which crucially rely on the self-monitoring of own attitudes and behaviors, might, in fact, be hard to observe or at least hard to induce in larger parts of the general population.

## Conclusion

Despite these caveats, we believe that the current correlational and experimental findings make an important contribution to social-psychological prejudice research in shedding light on the intra-individual cognitive dynamics that underlie the subtlety of explicit ethnic prejudice and have been largely abandoned within previous work. Proving the fruitfulness of applying a cognitive consistency perspective, this

research could reveal, based on two studies with large population samples, that own ethnically prejudicial beliefs indeed tend to be rather subtle to oneself, but that the degree to which this is true depends on egalitarian standards related to the condemnation of personal prejudice endorsement. Our findings show that individuals with high personal egalitarian standards (or for whom these are made salient) more strongly critically monitor their own opinions as potentially prejudiced and endorse prejudicial beliefs only to the extent that they are unaware of their xenophobic nature, thereby maintaining a consistency with these anti-prejudice standards. These results suggest that making people aware of the prejudicial content of their own opinions might not be an equally suitable means to combat explicit ethnic prejudice for all individuals, but can be an especially effective method to target people with highly egalitarian self-standards, since they are more unaware of the prejudicial content of their own views, in the first place.

## Notes

1. We use the terms ethnic prejudice and xenophobia here to refer to explicit negative attitudes directed at ethnic or national minority groups, and refer to single statements being reflective of such attitudes as ethnically prejudicial beliefs or statements.
2. It should be noted, that the current article does not target individuals' (un)awareness of implicit, automatic prejudicial biases they might exhibit (such as automatic stereotypical thoughts or negative affective or behavioral reactions), as for instance the bias awareness scale (Perry, Murphy, & Dovidio, 2015) and broadly also the should-would discrepancy questionnaire (e.g. Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith & Voils, 1998) are concerned with. In contrast to such approaches, the current article specifically examines the potential (un)awareness of individuals that their own *explicit* ethnically prejudicial opinions, i.e., beliefs they consciously endorse, are actually xenophobic and reflective of ethnic prejudice.
3. Note that with this conceptualization of external motivation to respond without prejudice as a standard that condones personal endorsement of ethnic prejudice we particularly refer to private settings, as it has been found that especially individuals with a high external (and low internal) motivation to be unprejudiced—which might be conceptualized as reflecting weakly self-determined anti-prejudice standards (see also Legault, Green-Demers, Grant, & Chung, 2007)—regulate ethnically prejudiced responses less strongly in private contexts (e.g., Butz & Plant, 2009). In public situations, we suppose that (primarily) externally motivated individuals might respond differently and potentially adjust their responses in line with social anti-prejudice norms,

because in that case the explicit labeling own beliefs as xenophobic or prejudicial in front of others would be inconsistent with their externally oriented anti-prejudice standards (see e.g., Butz and Plant, 2009, and Plant and Devine, 1998, for findings on the situational adjustment of prejudiced responses for primarily high-EM individuals).

4. For the sake of simplicity, we treat prejudice endorsement, prejudice awareness and endorsement of (anti-)egalitarian values as binary variables, when outlining our theoretical framework. However, we of course acknowledge and methodologically operationalize these constructs as continuous variables.
5. This study is based on a dual-frame (i.e., landline and mobile phone) sample. The landline sample was drawn based on the Gabler-Häder procedure (Häder & Gabler, 1998) that randomly generates and selects phone numbers. Random selection on the household level was ensured by choosing the adult who had the most recent birthday to answer the survey (Last-Birthday-Method). The mobile phone sample was drawn based on a similar procedure, however without the application of the Last-Birthday-Method, assuming that each mobile phone belongs to only one individual.
6. We deliberately used the term 'xenophobic' as an operationalization of prejudice awareness, since the term 'prejudiced' in everyday language means (at least in German) 'untrue', which would essentially refer to the same thing as asking people about their personal endorsement of different statements. The term 'xenophobic' rather implies that a statement is 'reflective of a negative attitude towards migrants', which is more suitable in the context of the current studies.
7. Although items were selected based on factor loadings indicated by Plant and Devine (1998), especially the internal reliability of the shortened EM scale was not fully satisfactory. However, since alpha scores are necessarily lower for shorter scales (e.g., Field, 2013) and research with a population sample was successfully conducted with only one of the two EM items we selected (Abrams, Houston, van de Vyver, & Vasiljevic, 2015), we considered these consistency scores sufficient for further analyses.
8. We thereby followed the chronological order of their assessment in the survey, although in general both ways of association between prejudice endorsement and prejudice awareness would be theoretically plausible. However, it was not possible and not the aim of the current analyses to establish a causal relationship between the two variables. The models reported in Table 1 do not substantially change when entering prejudice endorsement as the outcome and prejudice awareness as the predictor variable and are presented in Table A1 in the supplemental material.
9. Results in Model 8 (Table 1) remained uniform when also including PWE and its interaction term with prejudice endorsement in the final model.
10. Results for the reported models in Table 1 were robust when controlling for respondents' gender, age,

immigration background, education, and individual levels of social desirability.

11. In line with the rationale put forward in this article, the cognitive consistency mechanism assumed to underlie the hypothesized effect of the two experimental manipulations of egalitarian standards salience should be mainly applicable to individuals with high personal egalitarian standards, since only for these individuals, awareness of own prejudiced responses should represent a discrepancy with personal standards. Thus, only for these individuals should cognitive consistency, i.e., restoring, achieving or preserving consistency with own anti-prejudice standards, serve as a motivator to more strongly regulate and critically monitor prejudice-related responses in reaction to the two experimental tasks, as potentially reflected by the predicted stronger negative intra-individual association of prejudice endorsement and prejudice awareness. However, although the expected average treatment effects of the experimental manipulations attributable to such cognitive consistency mechanisms should be primarily driven by the response behavior of individuals with initially high personal egalitarian standards, it is nevertheless possible that the tasks employed for the two experimental conditions could also serve as a general prime of egalitarianism. As indicated by other research (e.g., Wyer, 2010), such a general egalitarian prime might also lead to compliance for high-prejudiced individuals or individuals with low personal egalitarian or anti-prejudice standards, prompting them to regulate and monitor prejudiced responses. From that perspective, also these individuals might exhibit a stronger negative intra-individual association of prejudice endorsement and prejudice awareness in reaction to the experimental tasks.
12. At the end of the survey, two items were administered asking about participants' general self-evaluations on in how far they reflect their own beliefs as potentially prejudiced. These items were not further analyzed within this article.
13. Within this pretest, participants ( $N = 174$ ) were asked to rate 10 randomly selected items out of 51 statements relating to migrants on a 1 (*not xenophobic at all*) to 5 (*very xenophobic*) scale. The items, ranging from neutral to very xenophobic statements, were either adapted from existing scales or self-developed.
14. The results were also uniform, when running the analyses based on participants from all three conditions, i.e., including also the two experimental groups. The models reported in Tables 3 and 4 do not substantially change when entering prejudice endorsement as the outcome and prejudice awareness as the predictor variable and are presented in Tables A4 and A5 in the [supplemental material](#).

## Acknowledgements

The authors wish to thank Antonio Joseph DelGrande, Lena Olive Röllicke, Markus Schopp, and Carolina Waldner for their assistance in conducting the studies and preparing this

manuscript, as well as Martin Kroh for his valuable comments on an earlier version of this article.

## Funding

This work was supported in part by a grant from Germany's Federal Government Commissioner for Migration, Refugees and Integration, and in part by a grant from the Hertie Foundation under grant number [P1140035]. We acknowledge support by the Open Access Publication Fund of Humboldt-Universität zu Berlin.

## References

- Abrams, D., Houston, D. M., van de Vyver, J., & Vasiljevic, M. (2015). Equality hypocrisy, inconsistency, and prejudice: The unequal application of the universal human right to equality. *Peace and Conflict: Journal of Peace Psychology*, 21(1), 28–46. doi:10.1037/pac0000084
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Butz, D. A., & Plant, E. A. (2009). Prejudice control and interracial relations: The role of motivation to respond without prejudice. *Journal of Personality*, 77(5), 1311–1341. doi:10.1111/j.1467-6494.2009.00583.x
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. doi:10.1037/0022-3514.42.1.116
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–378. doi:10.1037/0022-3514.82.3.359
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. doi:10.1037/0022-3514.44.1.113
- Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, 60(6), 817–830. doi:10.1037/0022-3514.60.6.817
- Doll, J., & Dick, M. (2000). Protestantische Ethik und humanitär-egalitäre Ethik: Studien zur Entwicklung zweier deutschsprachiger Ethikskalen [The Protestant ethic and the humanitarian-egalitarian ethic: Studies to construct two German language ethics scales]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 21(2), 138–151. doi:10.1024//0170-1789.21.2.138
- Fehr, J., & Sassenberg, K. (2010). Willing and able: How internal motivation and failure help to overcome prejudice. *Group Processes and Intergroup Relations*, 13(2), 167–181. doi:10.1177/1368430209343116
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row Peterson.
- Fetz, K., & Kroh, M. (2019). *Prejudice in disguise: Which features determine the subtlety of ethnically prejudicial statements?* Manuscript submitted for publication.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock'n'roll*. (4th ed.). London, UK: Sage.



- Forscher, P. S., Cox, W. T. L., Graetz, N., & Devine, P. G. (2015). The motivation to express prejudice. *Journal of Personality and Social Psychology*, 109(5), 791–812. doi:10.1037/pspi0000030
- Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social Cognition*, 30(6), 652–668. doi:10.1521/soco.2012.30.6.652
- Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative-propositional evaluation model. *Social and Personality Psychology Compass*, 8(8), 448–462. doi:10.1111/spc3.12124
- Gawronski, B., & Brannon, S. M. (2019). What is cognitive consistency, and why does it matter. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (pp. 91–116). Washington, DC: American Psychological Association.
- Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin*, 34(5), 648–665. doi:10.1177/0146167207313729
- Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, 40(4), 535–542. doi:10.1016/j.jesp.2003.10.005
- Häder, S., & Gabler, S. (1998). Ein neues Stichprobendesign für telefonische Umfragen in Deutschland [A new sampling design for telephone surveys in Germany]. In S. Gabler, S. Häder, & J. H. P. Hoffmeyer-Zlotnik (Eds.), *Telefonstichproben in Deutschland* (pp. 69–88). Opladen/Wiesbaden, Germany: Westdeutscher Verlag.
- Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, 23(2), 253–283. doi:10.1111/0162-895X.00281
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319–340. doi:10.1037/0033-295X.94.3.319
- Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., ... Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO(7) scale. *Journal of Personality and Social Psychology*, 109(6), 1003–1028. doi:10.1037/pspi0000033
- Ho, A. K., Sidanius, J., Pratto, F., Levin, S., Thomsen, L., Kteily, N., & Sheehy-Skeffington, J. (2012). Social dominance orientation: Revisiting the structure and function of a variable predicting social and political attitudes. *Personality and Social Psychology Bulletin*, 38(5), 583–606. doi:10.1177/0146167211432765
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, 55(6), 893–905. doi:10.1037/0022-3514.55.6.893
- Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, 40(3), 414–431. doi:10.1037/0022-3514.40.3.414
- Legault, L., Green-Demers, I., Grant, P., & Chung, J. (2007). On the self-regulation of implicit and explicit prejudice: A self-determination theory perspective. *Personality and Social Psychology Bulletin*, 33(5), 732–749. doi:10.1177/0146167206298564
- Levin, S., Sidanius, J., Rabinowitz, J. L., & Federico, C. (1998). Ethnic identity, legitimizing ideologies, and social status: A matter of ideological asymmetry. *Political Psychology*, 19(2), 373–404. doi:10.1111/0162-895X.00109
- Manganelli Rattazzi, A. M., & Volpato, C. (2003). Social desirability of subtle and blatant prejudice scales. *Psychological Reports*, 92(1), 241–250. doi:10.2466/PRO.92.1.241-250
- McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 91–126). San Diego, CA: Academic Press.
- McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, 25(4), 563–579. doi:10.1177/002200278102500401
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspective*, 2(4), 201–218. doi:10.1207/s15366359mea0204\_1
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112–117. doi:10.1111/j.1467-8721.2009.01619.x
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43. doi:10.1037/0022-3514.81.1.33
- Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, 65(3), 469–485. doi:10.1037/0022-3514.65.3.469
- Monteith, M. J. (1996). Affective reactions to prejudice-related discrepant responses: The impact of standard salience. *Personality and Social Psychology Bulletin*, 22(1), 48–59. doi:10.1177/0146167296221005
- Monteith, M. J., & Mark, A. Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. *European Review of Social Psychology*, 16(1), 113–154. doi:10.1080/10463280500229882
- Monteith, M. J., Mark, A. Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes and Intergroup Relations*, 13(2), 183–200. doi:10.1177/1368430209353633
- Monteith, M. J., & Voils, C. I. (1998). Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality and Social Psychology*, 75(4), 901–916. doi:10.1037/0022-3514.75.4.901
- Moskowitz, G. B. (2002). Preconscious effects of temporary goals on attention. *Journal of Experimental Social Psychology*, 38(4), 397–404. doi:10.1016/S0022-1031(02)00001-X
- Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, 47(1), 103–116. doi:10.1016/j.jesp.2010.08.014



- Moskowitz, G. B., Li, P., Ignarri, C., & Stone, J. (2011). Compensatory cognition associated with egalitarian goals. *Journal of Experimental Social Psychology*, 47(2), 365–370. doi:10.1016/j.jesp.2010.08.010
- Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual Review of Psychology*, 67(1), 363–385. doi:10.1146/annurev-psych-010213-115120
- Murray, K. E., & Marx, D. M. (2013). Attitudes toward unauthorized immigrants, authorized immigrants, and refugees. *Cultural Diversity and Ethnic Minority Psychology*, 19(3), 332–341. doi:10.1037/a0030812
- Paulus, C. (2009). *Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index* [The Saarbrueck Personality Questionnaire on Empathy: Psychometric evaluation of the German version of the Interpersonal Reactivity Index]. Retrieved from: <http://psydok.psycharchives.de/jspui/handle/20.500.11780/3343>
- Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, 3(3), 314–338. doi:10.1111/j.1751-9004.2009.00183
- Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology*, 61, 64–78. doi:10.1016/j.jesp.2015.06.007
- Pettigrew, T. F. (2018). The emergence of contextual social psychology. *Personality and Social Psychology Bulletin*, 44(7), 963–971. doi:10.1177/0146167218756033
- Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, 25(1), 57–75. doi:10.1002/ejsp.2420250106
- Pettigrew, T. F., & Meertens, R. W. (1996). The verzuiling puzzle: Understanding Dutch intergroup relations. *Current Psychology*, 15(1), 3–13. doi:10.1007/BF02686929
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832. doi:10.1037/0022-3514.75.3.811
- Plant, E. A., & Devine, P. G. (2009). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*, 96(3), 640–652. doi:10.1037/a0012960
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67(4), 741–763. doi:10.1037/0022-3514.67.4.741
- Schneider, S. L. (2007). Anti-immigrant attitudes in Europe: Outgroup size and perceived ethnic threat. *European Sociological Review*, 24(1), 53–67. doi:10.1093/esr/jcm034
- Schwarz, N., Strack, F., Hippler, H.-J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology*, 5(3), 193–212. doi:10.1002/acp.2350050304
- Sidanius, J., Pratto, F., & Bobo, L. (1996). Racism, conservatism, affirmative action, and intellectual sophistication: A matter of principled conservatism or group dominance? *Journal of Personality and Social Psychology*, 70(3), 476–490. doi:10.1037/0022-3514.70.3.476
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.
- StataCorp. (2017). *Stata Statistical Software (Release 15) [Computer software]*. College Station, TX: StataCorp LLC.
- Strabac, Z., & Listhaug, O. (2008). Anti-Muslim prejudice in Europe: A multilevel analysis of survey data from 30 countries. *Social Science Research*, 37(1), 268–286. doi:10.1016/j.ssresearch.2007.02.004
- Swim, J. K., Mallett, R., Russo-Devosa, Y., & Stangor, C. (2005). Judgments of sexism: A comparison of the subtlety of sexism measures and sources of variability in judgments of sexism. *Psychology of Women Quarterly*, 29(4), 406–411. doi:10.1111/j.1471-6402.2005.00240.x
- Voils, C. I., Ashburn-Nardo, L., & Monteith, M. J. (2002). Evidence of prejudice-related conflict and associated affect beyond the college setting. *Group Processes & Intergroup Relations*, 5(1), 19–33. doi:10.1177/1368430202005001536
- Wicklund, R. A., & Gollwitzer, P. M. (1981). Symbolic self-completion, attempted influence, and self-deprecation. *Basic and Applied Social Psychology*, 2(2), 89–114. doi:10.1207/s15324834baspp0202\_2
- Wicklund, R. A., & Gollwitzer, P. M. (1982). *Symbolic self-completion*. Hillsdale, NJ: Erlbaum Associates.
- Wyer, N. A. (2010). Salient egalitarian norms moderate activation of out-group approach and avoidance. *Group Processes and Intergroup Relations*, 13(2), 151–165. doi:10.1177/1368430209347326
- Zick, A. (1997). *Vorurteile und Rassismus: Eine sozialpsychologische Analyse [Prejudice and racism: A social-psychological analysis]*. Münster, Germany: Waxmann.
- Zick, A., Küpper, B., & Hövermann, A. (2011). *Intolerance, prejudice and discrimination: A European report*. Berlin, Germany: Friedrich-Ebert-Stiftung.
- Zuwerink, J. Z., Devine, P. G., Monteith, M. J., & Cook, D. A. (1996). Prejudice toward blacks: With and without compunction? *Basic and Applied Social Psychology*, 18(2), 131–150. doi:10.1207/s15324834baspp1802\_2